



The Milbank Memorial Fund
QUARTERLY

CONTENTS

	<i>Page</i>
IN THIS ISSUE	3
ACCURACY OF HEMOGLOBIN DETERMINATIONS ON FINGER-TIP BLOOD	<i>Dorothy G. Wichl</i> 5
VIBRATION SENSE AND FATIGUE	<i>Geoffrey Keighley, Ph.D.</i> 29
AN INSTRUMENT FOR MEASUREMENT OF VIBRATION SENSATION IN MAN	<i>Geoffrey Keighley, Ph.D.</i> 36
SOCIAL AND PSYCHOLOGICAL FACTORS AFFECTING FERTILITY V. THE SAMPLING PLAN, SELECTION, AND THE REPRESENTATIVE- NESS OF COUPLES IN THE INFLATED SAMPLE	<i>P. K. Whelpton and Clyde V. Kiser</i> 49



Vol. XXIV

JANUARY 1946

No. 1

Edited by the Technical Staff

Published quarterly by the MILBANK MEMORIAL FUND, 40 Wall Street,
New York 5, New York. Printed in the U. S. A. Subscription: \$1.00 a year.



T
health
they
the p
valu
glob
The
same
tip sp
tip b

Re
vibrat
"Vibr
whic
that
per
secre
4:00
for f
not s
grou
signi
that

Lo
devia

IN THIS ISSUE

THE determination of the amount of hemoglobin in blood is one of the most commonly used laboratory procedures in clinical and routine health examinations. Many methods are available and it is known that they vary greatly in the reliability of their results. For determinations by the photoelectric method, one of the most accurate methods, variations in values obtained are analyzed in the report entitled "Accuracy of Hemoglobin Determinations on Finger-Tip Blood," by Dorothy G. Wiehl. The error of the procedure found for repeated determinations on the same blood specimens and the sampling variation for independent fingertip specimens are discussed. Published data on the comparability of fingertip blood and venous blood are reviewed.

• • •

Results of an investigation of a possible relation between fatigue and vibratory sensation are reported by Dr. Geoffrey Keighley in the paper, "Vibration Sense and Fatigue." Two experiments were conducted in which vibration sense was measured in terms of the number of seconds that the subject could feel the vibration of a tuning fork of 128 cycles per second held between two fingers. Nine persons doing laboratory or secretarial work were tested on five or more days at 9:00 A.M. and after 4:00 P.M. In the other experiment, twenty-two men went without sleep for four successive days. The average threshold for vibratory sense was not significantly different in the morning and late afternoon for the first group, and the mean threshold for the second group did not change significantly after four days without sleep. Thus, there was no evidence that fatigue modified the vibratory sensation of these subjects.

• • •

Loss of vibratory sensation, or diminished sensitivity, is a neurological deviation which has been found fairly frequently among apparently

healthy persons. It has been suggested that vibratory sensation may be affected by nutritional deficiency. The usual test for vibration sense is the application of a tuning fork to fingers or toes of the subject, and the strength of the stimulus thus applied may vary considerably. A more controlled test in which an electrically driven vibrator is used and a quantitative measurement of the amplitude of the applied vibration is obtained is described by Dr. Geoffrey Keighley in the article entitled "An Instrument for Measurement of Vibration Sensation in Man." From tests on about 400 adults, it is shown that the strength of the stimulus required to produce a response increases as the frequency of vibrations per second increases. Also, the mean threshold for strength of stimulus was found to increase as age increased.

• • •

This issue continues the series of reports resulting from the work of the Committee on the Study of Social and Psychological Factors Affecting Fertility. This Committee's work has been sponsored by the Milbank Memorial Fund with grants from the Carnegie Corporation of New York. Two types of field work have been carried out in Indianapolis, the city chosen for study. First, a rapid "household survey" was made in which virtually all white families of Indianapolis were visited in order to locate couples meeting the qualifications for inclusion in the second type of investigation, an intensive study of social and psychological factors influencing size of family. The household survey provided the addresses of 2,589 couples appearing as eligible for the intensive study. Since budgetary considerations precluded intensive interviews with all of these couples, sampling was used. The sampling was done differentially by size of family, with the aim of securing a sufficient number of records for planned families of each size to permit separate analysis. The necessary adjustment of this sample to obtain properly weighted average rates regardless of family size was made by a conventional method of inflating the sample. The sampling plan and the representativeness attained are discussed in the article by P. K. Whelton and Clyde V. Kiser: Social and Psychological Factors Affecting Fertility. V. The Sampling Plan, Selection, and the Representativeness of Couples in the Inflated Sample.

ACCURACY OF HEMOGLOBIN DETERMINATIONS ON FINGER-TIP BLOOD¹

DOROTHY G. WIEHL

AN analysis of the sources of errors and their magnitude for hemoglobin determinations made by the photoelectric method is presented in this report. The hemoglobin estimates were part of serial examinations on a group of employed women. In order to evaluate fluctuations in the hemoglobin levels at different times, it is necessary to have a measure of the variation in estimated amounts of hemoglobin in the blood that is to be expected as a result of the experimental error for the method used in determining the hemoglobin values. Since many studies on hemoglobin are made which involve exact comparisons of values obtained at different times or under changed conditions the accuracy shown for the photoelectric method by this analysis is of interest.

The blood sample used for determining the amount of hemoglobin was taken by pricking the finger-tip. The finger-tip sample is easily and quickly obtained and, as a rule, the subject is less likely to object to having the finger pricked than to having a venipuncture sample taken. For surveys and in the routine practice of most physicians, a sample taken from the finger or ear-lobe is usually the preferred method of taking the blood specimen.

STUDIES COMPARING VENOUS AND CUTANEOUS BLOOD

The comparability of cutaneous or peripheral blood with venous blood in respect to hemoglobin content and the number of red blood cells has been investigated in a number of studies. These studies have been reviewed and if the data were published in sufficient detail, the findings are summarized in Table 1, and computations have been made to estimate the significance of differences shown for venous and cutaneous blood specimens. It was found that in the

¹ From the Milbank Memorial Fund.

Table 1. Summary of results of various studies comparing the hemoglobin content and the red blood cell counts for venous blood and cutaneous blood with new estimates¹ of the statistical significance of differences in reported values.

SOURCE OF BLOOD AND VALUES COMPARED	NUM- BER OF CASES	MEAN DIF- FER- ENCE	ST'D. ER- ROR OF MEAN	PROBA- BILITY OF CHANCE OC- CUR-	BASIS FOR ESTIMATED STANDARD ERROR OF MEAN
ADULTS					
<i>Hemoglobin;</i> Vein—Finger Andresen and Mugrage, 1938 (2)	30	— .06 gm.	±.061	>.30	Reported C. V. of 2.2% for 10 de- terminations on same sample
Rud, 1922 (6)	18	+ .33%	±.396	>.40	Differences between paired values
Vein—Ear Price-Jones, 1935 (3)	90	— 1.21%	±.168	<.001	Reported C. V. of 1.2% for values on 10 days for same person
Rud, 1922 (6)	27	+ .26%	±.554	>.60	Differences between paired values
Vein—Artery ² Gibbs, 1942 (4)	50	— .054 gm.	±.0107	<.001	Differences between paired values
<i>Red Blood Cells;</i> Vein—Finger Andresen, 1938 (2)		Millions			
	30	— .037	±.012	<.01	Reported C. V. of 1.3% for 10 de- terminations on same blood sample
Rud, 1922 (6)	18	+ .056	±.038	>.10	Differences between paired values
Bogendorfer, 1921 (7)	46	+ .518	±.067	<.001	Differences between paired values
Vein—Ear Price-Jones, 1935 (3)	96	— .078	±.0056	<.001	Reported C. V. of 0.7% for values on 10 days for same person
Rud, 1922 (6)	41	— .075	±.034	.02—.05	Differences between paired values
Duke, 1922 (8)	9	+ .082	±.047	>.10	Differences between paired values
Bing, 1919 (9)	10	— .320	±.208	>.10	Differences between paired values
CHILDREN AGED 2-14 YEARS					
<i>Hemoglobin;</i> Vein—Finger Andresen (2)	30	+ .03 gm.	±.055	>.50	Reported C. V. of 2.2%
Vein—Ear Rud (6)	4	0			
<i>Red Blood Cells;</i> Vein—Finger Andresen (2)	30	— .012	±.011	>.20	Reported C. V. of 1.3%
Vein—Ear Rud (6)	11	— .023	±.087	>.70	Differences between paired values

¹ See text, pp. 7 and 8.

² Blood from internal jugular vein and from the femoral, radial or brachial artery. Original data given in oxygen combining capacity.

original articles either no statistical test of significance for differences was applied to the data or the test used was not the most appropriate one for the problem.

If the results from samples of venous and of cutaneous blood were published for individuals, differences between the two values for each person were computed.* A frequency distribution of differences is obtained for which a standard deviation of the distribution and standard error of the mean of differences are computed in the usual way. Since it is desired to know whether there is any constant or systematic difference between venous and cutaneous blood, differences are computed with signs. Thus, for Table 1, the value for cutaneous blood was always subtracted from the value for venous blood and a minus difference indicates that cutaneous blood had the higher value. The mean of the differences computed in this way is the same as the difference between means for any set of original values. The reliability of this mean difference is tested only by the variation associated with the two types of blood samples for the same person and the variation in blood values for different individuals is eliminated since this variation is not related to the problem.

In two extensive statistical studies of venous and cutaneous blood, Andresen and Mugrage (2) and Price-Jones, Vaughan and Goddard (3), data for individuals were not published and the statistical significance of mean differences was measured in terms of the standard errors of the means derived from the distribution of values for different individuals and not, as described above, in terms of the variation between paired values for the same individual. Since hemoglobin values vary widely among individuals, the standard error of the difference in means for the observed values is considerably larger than the standard error of the mean of differences between paired values and means were found to be not significantly

* For examples of the use of differences between paired values and a discussion of their use in place of a comparison of the two distributions of values, see Snedecor (1).

different by this method which undoubtedly would be significantly different if judged by the variation between paired values. In both of these studies, some data are given on the accuracy of the hemoglobin method used and of the red cell counts and from this information a standard error of the mean difference has been computed which measures the probability that the observed mean difference would occur if the experimental error of the method were the only source of variation between the paired values. Thus, if the mean of the differences is significantly greater than is expected on the basis of error variation, it may be concluded that some other factor was operating, and in this case the factor presumably would be the source of the blood samples. If the standard deviation for experimental error of the method is based on a sufficient number of observations which include all sources of error which may affect the determinations on venous and cutaneous blood, the estimated standard error for a mean difference affords a valid measure for judging the observed difference. In these two studies, the coefficient of variation for the error of the method is given and is based on ten determinations which are too few observations to give a highly reliable estimate of the error. Furthermore, Andresen and Mugrage used ten determinations on the same blood sample and consequently their experimental error does not include any variation that might be associated with independent samples from the same person. Therefore, the estimated standard errors of mean differences* which would be expected from experimental error in the

* The standard error of the mean difference was computed from the coefficient of variation for experimental error as follows:

1. Coefficient of variation = $\frac{\text{S.D.} \times 100}{\text{Mean}}$; therefore, the standard deviation = $\frac{\text{C.V.} \times \text{Mean}}{100}$.

This standard deviation is the experimental error for a single determination. Andresen and Mugrage made duplicate determinations for each value and used the average; and the experimental error of this average value would be $\sqrt{\frac{\text{S.D.}^2}{2}}$.

2. For a series of differences between two values, each of which has a given experimental error (S.D.), the standard deviation = $\sqrt{2} (\text{S.D.})^2$.

3. The standard error of the mean of the differences is the standard deviation for the differences divided by \sqrt{N} and N is the number of pairs or number of differences.

method are somewhat too low for the Andresen and Mugrage data. From the study by Price-Jones, *et al.*, the coefficient of variation for determination on venous specimens taken from the same person on ten different days has been used for the experimental error.

From Table 1, it is evident that the difference between hemoglobin determinations and between red cell counts on venous and cutaneous blood is very small both for adults and for children aged 2 to 14 years. For hemoglobin, only the difference reported by Price-Jones between determinations on venous blood and blood from the ear-lobe is found to be significant,⁴ and the probability that the difference would occur as the result of accidental variation is less than one in a thousand. Although the hemoglobin values on venous blood determined by Price-Jones are significantly lower than those for blood from an ear-lobe, the difference was only 1.21 per cent hemoglobin, or about .17 gm., and for most clinical purposes so small an amount would not be important. Data from Gibbs, *et al.* (4), which compares hemoglobin in venous and arterial blood have been included in Table 1, and indicate a significant tendency for venous blood also to be very slightly lower in hemoglobin than blood from the femoral, radial or brachial artery. The difference is only .054 gm. and is of interest chiefly because it suggests that slightly higher hemoglobin values for blood from the ear-lobe and finger-tip may not be entirely due to irritation and congestion from the incision, as suggested by Drucker (5), or to other artificial causes.

Comparisons of red cell counts in venous and cutaneous blood are available from a larger number of studies and, in general, show results similar to those for hemoglobin. In the majority of studies, there is a tendency for the cell count to be somewhat lower for venous blood than for blood from a finger-tip or an ear-lobe, but the difference is not always statistically significant. For the Price-Jones data, the difference is very significant⁴ and the data from Andresen and

⁴In the original report, using the standard error of the difference in means for the distributions of values for individuals studied, the authors conclude that the difference is
(Continued on page 10)

Mugrage show a significant difference for adults but not for children.⁶ Values from Rud (6) also show a significantly lower cell count for venous blood for adults as compared with blood from the ear-lobe, but the counts on venous blood did not differ significantly from those on finger-tip blood. From one study, that of Bogendorfer and Nonnenbruch (7), a significantly higher cell count on venous blood samples was reported. Since Bogendorfer and Nonnenbruch found that this difference could be eliminated by putting the finger in hot water before taking the blood, it is possible that the puncture had not been sufficiently deep to establish a free flow of blood or that other artificial factors affected the findings. From the available data, it seems that the red cell count on venous blood may be very slightly lower than on that from the finger or ear-lobe. Inconsistent results and lack of significance in some studies could be explained as the result of the effect of such factors as the relatively large experimental error for red cell counts, the small number of cases compared, the technique of taking blood samples, and artificial and other factors affecting circulation of blood in the capillaries and arterioles.

Hemoglobin values and red blood cell counts seem to show a greater difference for venous and cutaneous blood in infants than in adults and older children. For thirty infants aged 1 to 19 months, Andresen and Mugrage obtained hemoglobin values on venous blood that averaged 0.55 gm. less than values on blood from the heel, and red cell counts were .051 million lower. For thirty infants one-half hour to nineteen days old, venous blood values averaged 0.77 gm. of hemoglobin and 0.264 million red cells lower. These not significant. If the experimental error used for the standard error of the mean of the differences in Table 1 were increased about three times for hemoglobin determinations and five times for red cell counts, the mean difference would still be significant.

⁶ Andresen and Mugrage also compared the volume of packed cells. For adults, the mean volume for venous blood was the same as the mean for finger blood; and for children, the mean volume for venous blood was slightly higher. Thus, the results for hemoglobin, number of cells, and volume of cells are not consistent. Furthermore, the validity of the significantly lower cell count on venous blood for adults is questionable since the reported experimental error by which the difference was tested probably is too low for error variation for independent samples.

differences would almost certainly be very significant statistically if tested by the method of differences between paired values. It is apparent that the difference was much larger for the very young infants. For six infants five to twenty-four days old, Haden and Neff (10) reported red cell counts for blood from the longitudinal sinus much lower than counts for blood from the heel in five cases (differences were 0.68 to 2.71 million), and higher by 0.11 million for one case. On the other hand, Lucas, *et al.* (11) reported higher hemoglobin values and red cell counts for blood from the longitudinal sinus than for unspecified peripheral blood for infants one to eight days old. Average values for sixty to 100 infants are given but comparisons are not made between averages for the same infants. Thus, the findings on infants have not been consistent, but the more controlled, recent study by Andresen and Mugrage strongly supports the view that in infants, especially the new-born, hemoglobin values and red cell counts are considerably lower for blood taken from the vein than for that from the heel.

In general, it may be concluded that, except for infants, blood from the finger or ear-lobe is not appreciably different from blood taken from a vein, although there is some evidence that both hemoglobin values and red blood cell counts tend to be slightly lower for venous blood. For most purposes,* cutaneous blood gives satisfactory results which are comparable with determinations on venous blood. However, since there may be a small, systematic difference, for special investigations in which small changes or differences are studied and tested statistically, values for venous blood and cutaneous blood should not be used interchangeably or considered to be identical. This precaution in making comparisons of values for blood from the two sources is necessary whether the difference is in part a true difference in venous and capillary or arteriolar blood

* Comparisons of venous and cutaneous blood have been for so-called healthy or normal persons. Duke and Stofer (8) found much higher red cell counts for blood from an ear-lobe than for venous blood in cases of pernicious anemia, but not in cases of secondary anemia. For hematological study of abnormal bloods, venous specimens seem preferable.

or is entirely due to the effects of the capillary technique, such as irritation, constriction, and congestion.

METHODS AND DATA COLLECTED

The hemoglobin values for finger-tip blood used in the following analysis were obtained by the Evelyn method (12) for photoelectric determination of oxyhemoglobin. A single Evelyn Photoelectric Colorimeter was used for all determinations. The instrument had been purchased in 1941, approximately a year before it was first used for the serial examinations of this study. The manufacturer's calibration and K_2 value were accepted. Any error in calibration would be constant for all values and would have no effect on this analysis of experimental error.

The procedure was to pierce the tip of a finger with a spring lancet and draw 20 cu. mm. of blood into a calibrated capillary pipet from which it was discharged into 10 cc. of distilled water in a colorimeter tube. The pipet was thoroughly rinsed by drawing the distilled water into the pipet about three times. After about ten minutes one drop of 28 per cent ammonia water was added and the tube was thoroughly shaken. The outside of the tube was carefully wiped before it was inserted in the colorimeter. The galvanometer was read to the nearest one-quarter unit of scale.

Two technicians made all the hemoglobin determinations on finger-tip blood samples. A single technician carried out the entire procedure for the hemoglobin determinations used in this analysis. One technician made all the determinations in a given examination period except for a few cases which have not been used in this report. Technician B made the determinations in two periods and Technician C in four periods. Both technicians had had considerable previous experience with this method of hemoglobin determination.

One technician (B), used the same capillary pipet for all blood specimens at both examinations and variation in determinations due to differences in accuracy of pipets is eliminated. The other tech-

nician (C), used a different pipet for the right and left-hand blood specimens and had new pipets for examinations in each period. In a given period, only two pipets were used except for an occasional sample taken with a third extra pipet. The two pipets were used at random for right and left specimens. Between specimens the pipet was thoroughly cleansed with distilled water and alcohol, then with acetone.

Hemoglobin determinations were made on a group of employed women who were healthy enough to be at work. Six examinations were made at about six-month intervals but many did not have all examinations. Except at the first examination, determinations were made for the right and left hand. These two independent hemoglobin values furnish the data for an estimate of the experimental error for the determinations on finger-tip blood.

DIFFERENCE BETWEEN TWO SAMPLES OF FINGER BLOOD

The mean difference between right and left-hand hemoglobin values for determinations made in each of five periods is shown in Table 2. These means are for differences with signs and the left-hand value was always subtracted from the right-hand value. A plus value, therefore, indicates that the right-hand value was higher.

Table 2. Difference between hemoglobin values for blood taken from a fingertip of the right and left hand.

TECHNICIAN AND EXAMINATION	NUMBER OF CASES	MEAN DIFFERENCE GMS. OF Hb. (RT.—LT.)	STANDARD ERROR OF MEAN DIFF.	STANDARD DEVIATION
TOTAL	391	+ .061	.013	.250
Technician C	261	+ .037	.015	.244
Examination 2	112	+ .016	.023	.241
Examination 4	95	+ .053	.027	.264
Examination 6	54	+ .051	.028	.208
Technician B	130	+ .115	.022	.256
Examination 3	85	+ .159	.030	.273
Examination 5	45	+ .030	.029	.197

In four of the five periods, the mean difference between right and left-hand values is very small, ranging from +.016 gm. to +.053 gm. These mean differences are not significant in a statistical sense except the highest value of .053 gm. which is just at the conventional line of significance and has a probability of occurring from chance variations of five in one hundred times. A much higher mean difference (+.159 gm.) is found for the two blood samples taken by Technician B in the third examination period and this mean is very significant. It will be noted that at each examination the sign of the mean difference is plus, and for the total cases for each technician the mean difference is significant. Apparently some factor operated to produce values that were slightly higher for the right hand. The cause of this bias is unknown. The most likely cause would seem to be some defect in technique. From the data available, the bias may be due either to a tendency for right-hand values to be too high or left-hand values to be too low.

As previously stated, Technician C used different pipets for the right and left-hand blood specimens and new pipets at each examination. The standard deviations for differences between the two specimens are about the same for both technicians and, therefore, there is no evidence that differences in pipets affected the variation between specimens taken by Technician C. Although the small mean difference between the right and left-hand values could have resulted from pipet differences, if the pipets were not used at random, as the technician believed, it is very unlikely that inaccuracies in the pipets would account for consistently higher right-hand values in several different examination periods.

For the relatively large mean difference for Technician B in the third examination period, there is reason to think that values for the left hand were too low.⁷ Since the same pipet was used for both hands and the right specimen was always taken first, it seems pos-

⁷The mean of hemoglobin values for all women was slightly lower in the third period than at any other examination. If the left values were too low, the effect would be to lower the mean for the group (See Table 4).

sible that at times enough moisture remained in the pipet after cleansing to have a slight effect on the volume of the left-hand blood specimen.* If this dilution factor did affect the left-hand values obtained by Technician B in the third examination, it may have been in addition to the undetermined factor which caused a slight bias in the values obtained by Technician C.

The very small but significant bias disclosed for these duplicate determinations is an example of systematic error which is apt to escape notice.

ACCIDENTAL ERROR OF ONE HEMOGLOBIN DETERMINATION

The observed difference between the two hemoglobin values for the same person at a given time is the result of the systematic bias plus the accidental error[†] for each of the determinations. The bias error is small and its effect on the variation between two samples may be eliminated by subtracting the mean difference for all examinations in a specific period from each difference between paired values for that period. This adjusted difference is equal to the deviation of each observed difference from the mean difference, and the standard deviation for these deviations is shown in Table 2. Thus, the standard deviation in Table 2 represents the variation between the two samples which resulted from accidental error for the two determinations.

On the assumption that each determination contributed one-half of the variation between samples which resulted from accidental error, the standard deviation for error variation in one determination is the standard deviation for accidental variation between samples divided by the $\sqrt{2}$. The standard deviation for error in

*In the re-examination of the group in December, 1945, Technician B took the left-hand blood specimen first and, as before, used the same pipet. For fifty-one paired values, the mean difference is $-0.055 \pm .023$, and the right-hand value is significantly lower than the left-hand value. This apparently confirms the hypothesis that, in spite of routine use of alcohol and acetone, there was a tendency for some moisture to adhere to the interior of the pipet.

[†]The term accidental error is used here to include all random or non-systematic variation associated with taking the blood samples and the technique of making a hemoglobin determination.

one hemoglobin determination is shown in Table 3 for each examination period, and the average standard error is given for each technician and for all determinations.²⁰

From the total experience the average standard error of a hemoglobin determination is .174 gm., that is, a hemoglobin value on finger-prick blood in this study was accurate within plus or minus .17 gm. in two out of three times and within .35 gm. in ninety-five out of one

hundred times so far as error from accidental sources in the technique and from blood sampling are concerned.

There was considerable variation in the errors estimated for the five different series of determinations. The standard errors for different periods ranged from .14 to .19 gm., as shown in Table 3. Technician B had both the highest and the lowest accidental error, and these standard deviations are significantly different; but the error for his total determinations was about equal to the standard error for all determinations by the other technician. The highest standard error for Technician C also differed significantly from his lowest error. Obviously, the error variation was not constant, even though estimated from a rather large number of determinations in each period. The combined experience probably affords the best estimate of the accidental error in this type of hemoglobin determination.

Nearly all hemoglobin values for this group of women were

²⁰ The average standard errors for each technician and for all five examination periods shown in Table 3, were computed as weighted averages of the variance for error in each period and are not exactly one-half the square of the standard deviations given in Table 2.

Table 3. Standard error for accidental variation of a single hemoglobin determination.

Technician and Examination	Number of Determinations	Standard Error Gms. of Hb.
TOTAL	782	.174
Technician C	522	.172
Examination 2	224	.171
Examination 4	190	.187
Examination 6	108	.147
Technician B	260	.176
Examination 3	170	.193
Examination 5	90	.139

between 11.50 gms. and 14.50 gms. and the mean was about 13 gms. There was no evidence of a consistent difference in the magnitude of the standard error for hemoglobin values at different levels within this range. The accidental error of .174 gm. is about 1.3 per cent of the mean hemoglobin level for this group, that is, the coefficient of variation for accidental error is 1.3 per cent. Since the estimated hemoglobin value may be either higher or lower than the true value, the true value represented by a determination will be within a range of four times the error in ninety-five out of a hundred times, and a 1.3 per cent error means that a hemoglobin estimate is unlikely to differ by more than 5.2 per cent from the true value.

SOURCES OF ACCIDENTAL ERRORS

The error in these duplicate, right and left-hand, determinations is the composite or net result of errors from several sources. In addition to variation due to blood sampling, the technician may make errors in measuring the sample or reading the galvanometer, undetected turbidity or imperfectly cleansed colorimeter tubes may cause error, and the performance of the photoelectric colorimeter may be a source of slight variation. The contribution of sampling variation for independent finger-tip blood specimens to the total accidental error has been estimated by comparing this error with that found when duplicate hemoglobin determinations were made by the same photoelectric method on two subsamples of a venous specimen.

The accidental error of a hemoglobin determination on a subsample of a blood specimen taken by venipuncture was found to be .145 gm. per 100 ml. of blood. This was estimated from duplicates on two hundred persons examined in a survey of nutritional status of high school students in New York City (13). Estimates of the error were made from data on two groups of one hundred persons examined at different times, and the two estimates were almost identical although different technicians made the determinations

and the work of several technicians was included in each series of one hundred. This accidental error of .145 gm. for the procedure of making a determination¹³ is somewhat lower than the error of .174 gm. found for finger-prick samples. Since hemoglobin values were determined by identical methods in the two studies, it is reasonable to conclude that the higher error in the values on finger blood was due to differences in the blood samples which are not present when two subsamples of blood are measured from a single venous specimen.

The difference between the variance¹⁴ for error of hemoglobin values on the finger-tip blood and that for venous blood gives a measure of the variance for the error associated with finger-prick blood sampling. This difference is as follows:

Variance for error of Hb. values on finger-tip blood	.030184
Variance for error of Hb. values on venous blood	.021077
Difference	.009107

The standard deviation for variation in finger-prick blood samples, as measured by these data, is $\sqrt{.009107}$ or 0.095 gm. Thus, if there were no procedural error in determining hemoglobin, a value derived from finger-prick blood could be expected to be accurate within plus or minus 0.19 gm. in ninety-five out of one hundred determinations. The sampling error for finger blood was less than the procedural error¹⁵ and the finger-prick samples apparently were

¹³ It is of interest to compare the procedural error estimated for this large series of determinations made by the Evelyn photometric method with the error estimated from data reported by Evelyn (12) as typical of the accuracy of the method. From duplicate values for ten blood samples reported, we have calculated a standard error for one determination of .087 gm. Although this error is appreciably smaller than the error of .145 gm. for our routine, survey duplicates, the method has proven highly accurate for routine work.

¹⁴ The variance for the error is the square of the standard deviation of error. When variation arises from the effects of several factors, the total variance is the sum of the variances for the variation contributed by the separate factors. Thus, if total variance is known, and variance for one factor also is known, the difference between these variances may be taken as a measure of the variance for other factors.

¹⁵ The small variation in these blood samples taken by puncturing the finger-tip is in agreement with the findings of Berkson, Magath, and Hurn (14) who studied errors in
(Continued on page 19)

affected very little by either natural or artificial causes.²⁴ At an average hemoglobin level of 13.0 gms., this sampling error is about 0.7 per cent.

Variation in samples of venous blood was studied by Walters (15) who withdrew ten samples of blood from the same venipuncture and varied the position of the needle between samples. A series of samples was taken from ten young men. For each sample, six readings were made by the Newcomer method in a colorimeter and averaged. The standard deviation for the variation among the ten average values for the same person ranged from .143 gm. to .419 gm.; the mean variation was .282 gm. and the coefficient of variation was 1.75 per cent. This variation for venous samples may have been significantly affected by the error of the average values used in spite of having made six determinations for each sample. Variation in the volume of packed cells also was reported for the same venous samples, and the mean coefficient of variation was 2.05 per cent. Since the procedural error for volume of cells is small, most of this 2 per cent variation would be caused by sampling variation.

In a study of diurnal variations of hemoglobin McCarthy and Van Slyke (16) estimated hemoglobin by the CO capacity method for venous specimens taken six times during the day from twenty-three subjects. The average change in values from 2: to 5: p.m. was $-.051 \pm .105$ volume per cent of CO capacity and therefore not significant.²⁵ If the variation between these twenty-three paired

erythrocyte counts and found that variation in counts for different finger-tip specimens from the same person was not significantly greater than the variation for subsamples of the same specimens. Since the procedural error for cell counts is large, a small but real sampling variation would be easily masked.

²⁴ Although the sampling variation is relatively small, it should not be neglected as a source of variation. For example, sampling variation as well as procedural error affects differences between venous and cutaneous blood. The reliability of any observed variation in hemoglobin at different times or under different conditions is affected by sampling error.

²⁵ The hours 2: and 5: p.m. were selected because the difference was less than that between any other two periods and the variation between samples is least affected by any real diurnal variability. A careful analysis of the data from McCarthy and Van Slyke has been published by Mole (17) and indicates a significant diurnal fluctuation from 9: to 11: p.m. but not from 9: to 5: p.m. From the six diurnal observations from 9: to 11: p.m., by

(Continued on page 20)

values is assumed to be the result of procedural error in each determination plus sampling variation, and the standard deviation for this total experimental error of one hemoglobin value is estimated in the same way as described for the paired values on the right and left hand, we obtain a standard error for one hemoglobin estimate of .36 volume per cent of CO or an error of 1.8 per cent. The CO capacity method has a very low procedural error¹⁶ and most of the variation of 1.8 per cent can be attributed to the blood samples. It is almost the same as that noted above for sampling variations in the study by Walters (15). In both studies, the variation indicated for venous blood specimens is greater than that estimated from our data for blood from a finger-tip.

Since it is common practice to make two determinations of hemoglobin and average them in order to obtain more accurate values, the amount of reduction in the experimental error obtained by averaging two determinations may be considered. If the determinations are made on two independent finger-tip blood specimens, both the error of sampling and of the method are reduced. The variance (standard error squared) for the average is one-half the variance for sampling error plus one-half the variance for procedural error or $.021007 + .009107$, and the standard error of the average is

$\frac{2}{2}$

$\sqrt{.015092}$ or .123 gm. as compared with .174 gm. for one determination on a finger-prick sample. But if two readings are made from the same specimen, variance for the average is one-half the variance for procedural error $(\frac{.021077}{2})$ plus variance for the

sampling error (.009107) and the standard error of the average is

analysis of variance Mole obtains a standard deviation of .46 volume CO for "uncontrolled error" after eliminating variance for average diurnal change and for individual averages. This standard deviation for "uncontrolled error" is affected to some extent by irregular diurnal variation as well as technical error and sampling variation.

¹⁶ McCarthy and Van Slyke (16) reported that the average difference between duplicates was .06 volume per cent of CO capacity but give no data on standard deviation or range for differences.

$\sqrt{.019646}$ or .140 gm. For most purposes, the slight reduction in error attained by puncturing two fingers to obtain independent samples does not seem worth while. Although the error is not greatly reduced by duplicate determinations, two estimates are so easily made on a photoelectric colorimeter that the greater accuracy is obtained with little effort and may be desirable at least for studies.

PHOTOELECTRIC METHOD COMPARED WITH OTHER METHODS

The foregoing estimates for variation in determinations of hemoglobin by the photoelectric method indicate a high degree of reproducibility. The method has been found highly accurate by others. Karr and Clark (18) compared hemoglobin values obtained by oxygen capacity method and by the photoelectric method using a Sheard-Sanford photelometer and concluded that "after an electric photometer is calibrated it gives more consistently accurate results than the oxygen capacity." However, under careful laboratory conditions, the Van Slyke oxygen capacity method has a somewhat lower error for duplicate determinations²¹ than the photelometric method but it is less suitable for surveys and routine work.

From a series of ten consecutive determinations on the same blood specimen made by the Hellige hemoglominometer, Andresen and Mugrage (2) reported a coefficient of variation of 2.2 per cent²² which may be compared with the procedural error for our routine determinations by the photelometric method of .145 gm. or about 1 per cent.

Most statistical data on accuracy of methods is found in studies comparing one method with another. In comparisons of hemoglobin estimates on subsamples of the same blood specimens by

²¹ The standard error of a Van Slyke oxygen determination estimated from duplicates on ten blood samples reported by Evelyn (12) was .07 gm. and the error from duplicates on seven samples made by an experienced laboratory technician for calibration of the photoelectric colorimeter used for the nutrition survey (13) determinations was .05 gm.

²² The error of a hemoglobin determination is commonly reported as a percentage or coefficient of variation. Our data and data from other studies which we have examined indicate that error variation is more constant in absolute amounts of hemoglobin than in percentage.

different methods, variation is the result of procedural error plus any bias or systematic difference in the measurement of hemoglobin. Therefore, the differences between estimates by different methods are not comparable with the differences for repeated determinations by the same method which measure only procedural error, but with adjustment of differences between two methods for any average difference, the variability due to a systematic factor, such as calibration of instruments, is eliminated and the remaining variation is largely that associated with the method. Karr and Clark (18) studied the accuracy of a number of methods in general use and had laboratory technicians and physicians make hemoglobin determinations on subsamples of the same blood specimen by various methods. They reported the per cent of determinations which differed by not more than .5 gm. and not more than 1.0 gm. from the value obtained by the photelometric method. After corrections for systematic bias due to calibration of instruments or personal bias in matching color, the best results were obtained by one technician using a Hellige-wedge type colorimeter who had 83 per cent of his determinations within .5 gm. and 93 per cent within 1.0 gm. of the standard value, and by one technician using a Haden-Hausser colorimeter who had 80 per cent of his determinations within .5 gm. and 98 per cent within 1.0 gm. of the standard. On the basis of our standard error of .145 gm. for each value, differences between two photelometric determinations would have a standard deviation of .205 gm. and 98 per cent would be expected to differ by not more than .5 gm. and none would differ by more than 1 gm. Each technician made hemoglobin estimations on about forty bloods, and allowing for the statistical reliability of their percentages, these two technicians obtained results only a little less accurate than might occur for photelometric determinations. But for four other technicians using the Haden-Hausser colorimeter, only 45 to 55 per cent of their determinations differed by not more than .5 gm. Two technicians using Sahli instruments had 46 and 54 per cent of the

values within .5 gm. of the standard; one technician using a Bausch and Lomb Newcomer instrument had 49 per cent of his determinations within .5 gm. of the standard. It is apparent that all of these methods furnished very unreliable estimates of the amount of hemoglobin, and if no correction in results is made for systematic bias the magnitude of the differences is much increased for most of the technicians. Results for the physicians' readings were less accurate than for technicians.

PERIODIC VARIATION IN DETERMINATIONS

The accidental error measurable from duplicate values does not reveal inaccuracies in the determinations which may result from some constant source of error. Inaccurate calibration of an instrument will give results which are consistently too high or too low by a constant amount, and an inaccurately calibrated pipet will have the same effect. These are systematic errors which must be guarded against and can be eliminated. But other systematic errors which may affect the determinations are more difficult to detect and unless a careful study is made the presence of bias or systematic error remains unknown. The personal equation is an important factor in methods which require the technician to match colors, and Karr and Clark (18) have shown that many persons tend to consistently underestimate or overestimate hemoglobin when matching color standards. The photoelectric method eliminates this personal factor. Evidence of other systematic bias was found in the determinations by the two technicians who examined the same group of women several times.

Comparison of the mean hemoglobin levels at six different periods for this group of women which are given in Table 4 shows that relatively high values were obtained at two examination periods. The higher values were for the second and sixth examinations, both by Technician C. In the first period, only one determination was made for each person and for the other five periods an average of the right and left-hand determinations was used. Not

EXAMINATION PERIOD	TECHNICIAN	NUMBER OF WOMEN ¹	MEAN GMS. PER 100 ML.	NUMBER OF WOMEN ²	MEAN GMS. PER 100 ML.
I. Oct., 1942	C	63	12.94	36	12.90
II. July, 1943	C	63	13.44	36	13.38
III. Nov., 1943	B	63	12.85	36	12.83
IV. May, 1944	C	63	13.03	36	13.20
V. Nov., 1944	B	38	12.92		
VI. May, 1945	C	43	13.58	36	13.57

¹ Same 63 women for periods I.-IV.

² Same 36 women for all five periods.

Table 4. Mean hemoglobin values for a group of women examined at approximately six-month intervals.

all women were examined at every period, and in the fifth and sixth periods both technicians made some hemoglobin determinations. Furthermore, women are excluded for whom the hemoglobin value was less than 11.0 gms. at any examination. Therefore, in Table 4, average hemoglobin values are shown for sixty-three women examined at each of the first four periods and for all those examined by a given technician in the fifth and sixth periods. It is apparent that the group mean values for the two periods in which Technician B made the determinations are very similar, and that the group mean values for two of the four periods in which Technician C made the determinations closely approximate the means for Technician B's values. But in two periods Technician C obtained values which were, on the average, higher than those in the other four periods. There was no consistent trend in the means and the two high periods were about two years apart³ with the lower means prevailing for three examinations in between them. The conclusion seems justified, though not definitely proved, that some systematic overestimate of hemoglobin occurred in the second and sixth periods.

³ The second examination (high mean level) was in July 1943 and the sixth examination (high mean level) in May 1945. Thus, one was a summer period and one a late spring period. The fourth examination also was in the spring, and the other three were in October, November, or December. It is possible that there was some seasonal variation, but it does not seem likely that an average decrease in hemoglobin of .59 gm. occurred between July and November.

SOURCE OF VARIATION	DEGREES OF FREEDOM	VARIANCE	RATIO	PROBABILITY
Individual Differences	35	1.7786		
Period Differences	4	3.5370	8.95	<.001
Remainder—Uncontrolled	140	0.3950		

Table 5. Analysis of variance of hemoglobin values for thirty-six women at five different periods.

If the fifth examination period is excluded, hemoglobin values are available for thirty-six women for the other five periods. Average hemoglobin values for these thirty-six women are given in Table 4, and an analysis of the statistical significance of the variation among average values for the five periods is given in Table 5. The variation among mean hemoglobin values for the different periods is very significant, that is, it is much greater than would be expected as the result of random variability due to accidental error and other influences on hemoglobin fluctuation.

For the thirty-six women, the average hemoglobin value in the fourth examination period is significantly higher³⁰ than that for the first and for the third periods, and not significantly lower than the average for the second period. Thus, for this smaller group of women, the determinations by Technician C are relatively high in three of the four periods in which he did the examinations. The mean value for the thirty-six women in the fourth period is somewhat higher than the mean for the larger group of sixty-three women and it seems probable that the tendency to obtain higher estimates of hemoglobin was operating to some extent but not as consistently as in the second and sixth periods.

Evidence that the performance of the photoelectric colorimeter was not a factor in obtaining higher hemoglobin values in some periods is obtained from ten determinations made by Technician C

³⁰ The standard error of the difference between any two means is $\pm .148$ gm. This is obtained from the uncontrolled variance in Table 5 and is $\sqrt{\frac{2 \times 0.3950}{36}}$.

in the fifth examination period when Technician B made all other determinations. For the ten women examined by Technician C, the averages of hemoglobin values in the fourth and fifth periods were 12.43 gms. and 13.16 gms., respectively, indicating a significant increase in hemoglobin at the later period. But determinations by Technician B show no similar tendency to be higher in the fifth period. This finding points strongly to some shift in the technique of Technician C. The most probable explanation seems to be that he had a tendency to measure the sample somewhat generously and possibly also to read the colorimeter scale with a slight bias.

It is of interest and important that the personal factor in the technique of Technician C was not constant throughout his four examination periods. He had both high and low periods and these were not associated with any change in the accuracy of his determination as revealed by variation between right and left-hand values. Apparently any shift in his attitude or standards for measurements was fairly constant throughout a specific examination period.

Two troublesome problems in laboratory data are illustrated by these findings for hemoglobin values at six different examination periods. These are (1) using the identical technique and equipment, technicians may obtain significantly different results due to certain personal bias or attitudes; and (2) the same technician may change with respect to these personal factors over a period of time. Only the most careful and frequent evaluation of the work of technicians will reveal the presence of systematic bias due to the personal equation or bias from other sources. Regardless of the high degree of accuracy of the method used, such variations in results may occur. If undiscovered, a systematic bias in data easily may lead to erroneous conclusions.

SUMMARY

Sources of variation in estimates of hemoglobin content of blood made by the Evelyn photoelectric method are discussed.

From duplicate determinations on venous blood specimens from

200 subjects examined in a survey, the standard error of one determination is estimated as $\pm .145$ gm. of hemoglobin per 100 ml. This is a procedural error only.

For 391 examinations in which two independent determinations were made using blood from a finger of the right and of the left hand, the standard error of one determination is estimated as $\pm .174$ gm. This error includes the procedural error and blood sampling variation. On the assumption that the procedural error was equal in the two studies, the standard error for sampling variation is $\pm .095$ gm. for finger-tip blood specimens.

An analysis of differences between hemoglobin values obtained on the right and left-hand blood samples indicated that values for the right-hand specimens, which were always taken first, were on the average slightly but significantly higher than those for the left hand. This difference could not be explained.

The determinations on finger-tip blood were from serial examinations made at approximately six-month intervals on the same group of women. One technician made the determinations in four periods and a second technician made them in two periods. Comparison of the average hemoglobin levels for the six examination periods indicates a significant variation in the average levels which seemed to be the result of some technical variation. A possible explanation is that one technician had a tendency at times to measure the blood samples slightly generously.

REFERENCES

1. Snedecor, George W.: *STATISTICAL METHODS*. Iowa State College Press, Ames, Iowa, 1940, chapter 2, p. 74.
2. Andresen, Marjorie I. and Mugrage, Edward R.: Venous and Peripheral Red Blood Cell Values. *American Journal of Clinical Pathology*, 1938, 8, pp. 46-51.
3. Price-Jones, C.; Vaughan, Janet M.; and Goddard, Helen: Hematological Standards of Healthy Persons. *Journal of Pathology and Bacteriology*, May 1935, 40, pp. 502-519.
4. Gibbs, E. L.; Lennox, G. W.; Nims, L. F.; and Gibbs, F. A.: Arterial and Cerebral Venous Blood. *Journal of Biological Chemistry*, 1942, 144, pp. 325-332.
5. Drucker, Paul: Investigations on the Normal Values for the Hemoglobin and Cell Volume in the Small Child. *Acta Paediatrica*, 1923, iii, pp. 1-39.

6. Rud, Einer J.: Le Nombre des Globules Rouge chez les Sujets Normaux et leur Variations dans les Diverses Conditions Physiologique. *Acta Medica Scandinavica*, 1922-1923, 57, p. 168.
7. Bogendorfer and Nonnenbruch: Vergleichende Bestimmung der Blutkörperschenzahl, des Serumweisses und Serumkochsalzes in Venen- und Kapillarblut. *Deutsches Archiv für klinische Medizin*, September 1920, 133, pp. 389-396, Leipzig.
8. Duke, W. W. and Stofer, D. D.: A Comparison of Capillary and Venous Blood in Pernicious Anemia. *Archives of Internal Medicine*, 1922, 30, p. 94.
9. Bing, H. I.: Antallet af rode Blodlegemer i de forskellige Aldre og under forskellige Forhold. *Ugeskrift for Laeger*, September 18, 1919, 812, No. 38, pp. 1483-1489.
10. Haden, R. L. and Neff, F. C.: The Volume Index and Color Index of the Red Blood Corpuscles in Newborn Infants. *American Journal of Diseases of Children*, 1924, 28, pp. 458-463.
11. Lucas, W. P.; Dearing, B. F.; Hoobler, H. R.; Cox, Anita; Jones, M. R.; and Smyth, F. S.: Blood Studies in the New-Born. *American Journal of Diseases of Children*, December 1921, 22, pp. 525-559.
12. Evelyn, K. A.: A Stabilized Photoelectric Colorimeter with Light Filters. *Journal of Biological Chemistry*, August 1936, 115, p. 63.
13. Kruse, H. D.; Palmer, C. E.; Schmidt, W.; and Wiehl, Dorothy G.: Medical Evaluation of Nutritional Status. I. Methods Used in a Survey of High School Students. *The Milbank Memorial Fund Quarterly*, July 1940, xviii, No. 3, pp. 257-298.
14. Berkson, Joseph; Magath, Thomas B.; and Hurn, Margaret: Laboratory Standards in Relation to Chance Fluctuations of the Erythrocyte Count as Estimated with the Hemocytometer. *Journal of the American Statistical Association*, June 1935, xxx, No. 190, pp. 414-426.
15. Walters, Orville S.: Variation of Erythrocytes, Hemoglobin and Packed Cell Volume in Immediately Consecutive Samples of Venous Blood. *American Journal of Physiology*, 1934-1935, 110, pp. 37-41.
16. McCarthy, Earl F. and Van Slyke, Donald D.: Diurnal Variations of Hemoglobin in the Blood of Normal Men. *Journal of Biological Chemistry*, May 1939, 128, pp. 567-572.
17. Mole, R. H.: Diurnal and Sampling Variations in the Determination of Haemoglobin. *Journal of Physiology*, June 1945, 104, No. 1, pp. 1-5, London.
18. Karr, W. G. and Clark, J. H.: Comparison of Various Hemoglobin Methods as Performed in Hospitals and Physicians' Laboratories. *American Journal of Clinical Pathology* (Technical Section), September 1941, 5, pp. 127-147.

VIBRATION SENSE AND FATIGUE

GEOFFREY KEIGHLEY, PH.D.¹

IN the performance of many tasks, fatigue occurs although the amount of muscular effort is small. General fatigue of this character is attributed to fatigue of the central nervous system (1). It is difficult to evaluate. Examination of the responses to various stimuli is one of the ways which have been tried. It is not inconceivable that somewhere in the chain of functions from sense organ to response there is one which would be sensitive to fatigue, causing a change of the threshold, probably raising it. Vibration sensitivity is known to be altered by cold, (2,3) by increasing age (2,4,5), and by severe vitamin deficiencies (6,7). Roth (8) has said that during general fatigue the vibration sense diminished, and suggested that its measurement might be of importance in aviation medicine for the estimation of the fatigue level in flight personnel. He gives no figures showing such a decrease. The experiments reported in this paper were performed to test Roth's suggestion that vibration sensibility is diminished during fatigue, and to examine the simple quantitative method which he has devised.

The method which Roth described is not only simple but, as shown below, has a relatively small inherent error. He uses a 128 cycles per second (C.P.S.) tuning fork with a cross bar in the stem. To start it with the same amplitude of vibration each time, he strikes it against his hand with enough force to cause the two ends to click together. He suspends the fork between two of the subject's fingers; thus, hanging freely by the cross bar, the vibration is applied under constant pressure. He takes the time from the moment he strikes the fork until the subject no longer feels the vibration. The elapsed time is the threshold. With the standard fork, Roth found the threshold to be thirty-five seconds for normal young

¹ From the William G. Kerckhoff Laboratories of the Biological Sciences, California Institute of Technology, Pasadena.

people, with a range of three or four seconds. The thresholds are lower in other parts of the body. There is a decrease with age, and if the fingers are cold.

The pressure of the freely hanging fork is uniform. The time can be measured accurately with a stop-watch, started when the fork is struck against the hand, and held in the hand which is struck. There is the question of how uniformly the fork is started with this improved method. Tests of this uniformity were performed only after extensive practice had perfected technique.

Strain gauges were cemented to the outer sides of both tines of the fork near the stem (10). The wires of the gauges, running longitudinally, were alternately stretched and compressed as the fork vibrated, thus changing their resistance. The changes were recorded photographically with an oscillograph. The width of the trace on the record can be measured readily; it is proportional to the amplitude of vibration.

Several records were made. The beginnings of each trace varied somewhat, showing that even though the greatest care was taken in striking, the fork did not start with uniform strength. Roth recommends striking the mid-third of one prong against the hypothenar eminence with sufficient force to click the ends together. This method gave variations in amplitude of the order of ± 9.0 per cent. If, instead of the mid-third the end of one prong is similarly struck, although the fork starts with somewhat less amplitude, the variations are of the order of only ± 3.5 per cent. Consequently, in all the experiments reported further on, the fork was started by striking the end. No more force should be used than is enough to cause the easily heard click.

Variations in amplitude at the start mean that there are differences in the length of time the fork takes to drop to a level at which its vibration could no longer be perceived. It is necessary to convert the variations in amplitude into differences of time. The strain gauges cause some damping. The less damped vibrations, without

the gauges, were measured with a microphone, amplifier, and output meter, by standing the vibrating fork upright on its stem on the back of the microphone, and taking readings of the output meter every five seconds. This is less accurate than using the strain gauges, but the mean of many runs gave a curve of the decay of amplitude with time which corresponds closely with the rate of decay when the fork is hanging from the fingers. Differences of amplitude between starts of ± 3.5 per cent, from the strain gauge records, projected onto this curve, show differences of ± 0.5 second in length of time taken to drop to a threshold amplitude.

This method of using a tuning fork is therefore sufficiently reliable and consistent to be used in making quantitative threshold measurements.

One of these 128 C.P.S. forks was used on three people to find the thresholds in the first two fingers, of right and left hands. Thresholds in both hands were determined once a day. Table 1 shows the comparisons; thresholds are given as the length of time in seconds the subjects felt the vibration. In one subject the mean thresholds of the two hands were equal, in the other two cases the differences were small, and not statistically significant.

The effects of fatigue were sought in the morning and afternoon

Table 1. Vibration threshold to tuning fork 128 cycles per second measured in seconds from time of striking to absence of sensation. Comparison of right and left hands.

SUBJECT	NO. OF TIMES TESTED	THRESHOLDS IN SECONDS				DIFF. OF MEANS	
		RIGHT HAND		LEFT HAND			
		MEAN	STANDARD ERROR OF MEAN	MEAN	STANDARD ERROR OF MEAN		
J. J.	14	26.1	± 0.7	26.1	± 0.7	0.0	
G. H.	19	36.9	± 0.9	37.0	± 0.7	0.1	
J. D.	18	33.1	± 1.0	35.1	± 1.0	2.0	

thresholds of a group of nine people (one male, J.D.). All were doing laboratory or secretarial work. Tests were made on the right hand only, twice a day for five or more days. Thresholds, from three or more trials with the fork hanging on the first two fingers, were measured in the morning starting at 9: a.m., and in the afternoon after 4: p.m. They are given in Table 2; all morning and afternoon thresholds are averaged for each subject.

If vibration thresholds were affected by fatigue, one might expect them to be consistently lower or higher at the end of a day's activities. They were not in the subjects tested. The differences, in seconds, between each subject's mean morning and afternoon thresholds, (D of the table), are not large. A ratio of the difference to the standard error of the difference (D/σ_D) of 2 or greater is commonly taken as significant. All the ratios are smaller than this. The changes are not even all in the same direction, eight out of the nine are positive, one negative.

There was no correlation between the results of the tests and subjective expression of fatigue which were volunteered or given in reply to occasional questions. It may be objected that, in spite of these subjective symptoms, the work these people were doing was

Table 2. Comparison of morning and afternoon vibration thresholds to tuning fork at 128 cycles per second, of nine people.

SUBJECT	NO. OF DAYS	MEAN THRESHOLD—SECONDS		D ¹	$\frac{D^2}{\sigma_D}$
		A. M.	P. M.		
F. A.	21	34.3	35.0	+0.7	0.7
M. K.	21	36.4	36.9	+0.5	0.5
V. B.	14	22.2	23.3	+1.1	1.3
J. W.	10	33.8	35.3	+1.5	0.6
H. R.	10	38.5	39.8	+1.3	0.8
G. H.	8	36.0	37.1	+1.1	0.6
J. D.	8	32.5	34.4	+1.9	1.2
J. J.	6	27.2	25.8	-1.4	0.9
M. E.	5	34.6	35.6	+1.0	0.4

¹ Difference between the mean thresholds.

² Ratio of the difference to the standard error of the difference.

not very hard, that they were accustomed to it, and that only moderate fatigue might be expected.

The results of a different experiment, summarized in Table 3, should meet these objections. The twenty-two men in this group went without sleep for four successive days. On two days before the experiment, and on each of the days without sleep they were tested each evening with the tuning fork. It was not practicable to repeat the test at the end, after they had had a night's sleep. The behavior and appearance of these men showed that they were under increasing stress, yet their vibration thresholds were not markedly changed. Their thresholds on the days without sleep differed from those on the days when they had had sleep, but the ratios of the differences to the standard errors of the differences are all less than 2, so the differences are not significant. Even though all the changes are positive they are too small to show any trend which can be associated with the increasing fatigue and strain.

In two experiments then, the first involving the fatigue brought on by the ordinary activities of a day's work, the second the cumulative fatigue of four days without sleep, the average thresholds before and after the fatiguing experience do not show any significant dif-

Table 3. Vibration thresholds to tuning fork at 128 cycles per second before and during four days without sleep for twenty-two men.

DAY	MEAN THRESHOLD SECONDS	D ¹	D ² — σ_D
Five Days Before Start	27.9		
Day Experiment Started	28.3		
After the Following Number of Days Without Sleep			
1	29.8	+1.9	1.2
2	29.4	+1.5	1.0
3	28.2	+0.3	0.2
4	30.6	+2.7	1.9

¹ Difference of the mean threshold of each experimental day from the mean five days before of 27.9 seconds. If taken from the mean of 28.3 seconds the differences and the ratios D/ σ_D are smaller.

² Ratio of the difference to the standard error of the difference.

ferences. Nor are individual changes reliable. The nine people in Table 2 were tested a total of 103 times; and their afternoon compared to their morning thresholds, were higher forty-eight times, lower forty-two times, and unchanged thirteen times; each of the nine had about an equal number of higher and lower afternoon thresholds. There was a similar random distribution of individual daily changes among the men who went four days without sleep.

These measurements at the fingers, made at 128 C.P.S., on two groups of people subjected to widely different degrees of stress, show that their vibration thresholds were not changed by the fatigue arising from the stress. Single day-by-day thresholds are unreliable for they may be raised or lowered. If the mean thresholds are taken of several people, or of one person over several days, the variations are less, though they may still not be consistently in one direction (*see Table 2*) and the differences are small and not significant (*Tables 2 and 3*).

SUMMARY

The simple, quantitative method described by Roth for measuring vibration sense at one frequency has been examined for its inherent error, which is found to be relatively small.

This simple method was used on two groups of people undergoing widely different degrees of stress. No changes in vibration thresholds could be found as the result of these differently fatiguing experiences.

REFERENCES

1. Simonson, Ernst: Industrial Physiology. *Annual Review of Physiology*, 1944, 6, pp. 543.
2. Keighley, Geoffrey: An Instrument for Measurement of Vibration Sensation in Man. *The Milbank Memorial Fund Quarterly*, January 1946, xxiv, No. 1, p. 36.
3. Weitz, Joseph: Vibratory Sensitivity as a Function of Skin Temperature. *Journal of Experimental Psychology*, 1941, 28, pp. 21.
4. Pearson, Gerald H. J.: Effect of Age on Vibratory Sensibility. *Archives of Neurology and Psychiatry*, 1928, 20, pp. 482.

5. Newman, Henry W. and Corbin, Kendall B.: Quantitative Determination of Vibratory Sensibility. *Proceedings of the Society for Experimental Biology and Medicine*, 1936, 35, pp. 273.

6. Fox, James C. Jr., and Klemperer, Wolfgang W.: Vibratory Sensibility, A Quantitative Study of Its Thresholds in Nervous Disorders. *Archives of Neurology and Psychiatry*, 1942, 48, pp. 622.

7. Shulack, Norman R. and Peters, Michael: Spinal Cord Changes in Avitaminosis. *Bulletin of the U. S. Army Medical Department*, 1945, No. 88, pp. 118.

8. Roth, Aaron: Measurement of Vibration Sense; A Method and a Means. *War Medicine*, 1943, 4, pp. 280.

9. Hathaway, Claude M.: Strain Analyzing and Recording Instruments. *Electronic Industries*, 1945, 4, pp. 74.

AN INSTRUMENT FOR MEASUREMENT OF VIBRATION SENSATION IN MAN¹

GEOFFREY KEIGHLEY, PH.D.²

M EASUREMENT of vibration sensation is commonly part of a clinical neurological examination, especially where particular attention is paid to possible nutritional disease. The method much used is application of a tuning fork to fingers and toes. In a nutrition survey which will be reported later (1) it was felt that a more quantitative method was desirable, one which would give thresholds, as amplitudes of the applied vibration, over a range of frequencies. An instrument for this purpose was devised and found to be useful in an examination of the thresholds, at the toes, of over 400 cases. This paper gives a description of that instrument, which is an electrically driven vibrator with amplitude and frequency independently controlled, together with some of the results obtained from its use. Similar apparatus has been described previously (2,3). The equipment may be assembled almost wholly from available, commercial, components.

A variable frequency audio oscillator generated the needed frequencies, the output from the oscillator was fed into an amplifier, the amplifier drove the vibrator unit which transformed the electrical oscillations into mechanical movements. A voltmeter was used to measure the strength of signal supplied to the vibrator unit and a cathode-ray oscilloscope was used as a check when necessary on the frequency of the oscillations by comparison with the frequency of the 115-volt 60 cycle supply to the apparatus.

The variable frequency oscillator was of the type in which the frequency is determined by a resistance-capacity network (4). This type of oscillator is very stable, the warming-up period is short,

¹ Abstracted from a thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, California Institute of Technology, 1944.

² From the William G. Kerckhoff Laboratories of the Biological Sciences, California Institute of Technology, Pasadena.

and there is little subsequent frequency drift. The frequency at which it operated was set by a potentiometer with a dial calibrated in terms of frequency. It provided, within limits, any possible frequency lying in its range of 20 to 20,000 cycles per second (C.P.S.), and from this wide range only certain frequencies were chosen. The ones used, from 50 C.P.S. up, are included in Table 1.

The amplifier was one covering the audio range, a Clarion Model A3. A volume control was used to regulate the output to the vibrator unit, and so the strength of the stimulus applied to the subject. The magnitude of the stimulus was recorded with the aid of a voltmeter in the output circuit of the amplifier. The voltage readings so obtained are proportional to the amplitude of the applied vibration. The sensitivity of the voltmeter was somewhat greater than necessary. For this reason and to make subsequent calculations simpler, the recorded voltages were grouped and converted into a series of arbitrary numbers and fractions ranging from 0, no movement of the vibrator, to 7, corresponding to the maximum stimulus which the apparatus could deliver. These numbers of course are also proportional to the amplitude of the applied vibration. The nature of the arrangement was such that each number represented a voltage and amplitude of vibration twice the size of the number below it, thus the amplitude 2 is twice as big as amplitude 1, 3 twice as big as 2, and so forth. The justification for using arbitrary numbers to record the magnitude of the vibration comes from the fact that even with the same apparatus used at the same level of input

Table 1. Amplitude of vibration at different frequencies with needle in place on a toe and constant input to vibrator.

	FREQUENCY, CYCLES PER SECOND												
	50	100	150	200	250	300	350	400	500	600	800	1,000	1,500
Amplitude (Relative)	14	14	8	8	7	8	8	8	9	8	5	4	3

to the vibrator, the response of the same subject will vary according to the effective area of the vibrating member applied to the skin, and the pressure with which it is applied; there are no standards for these two dimensions. These dimensions however were kept constant (*see below*).

The vibrator unit consisted of an Astatic M 41 magnetic cutting head which is designed for making phonograph records. The usual cutting stylus was replaced by a needle made of music wire 0.036" in diameter and 1" long with a flat brass button 0.08" in diameter soldered to the end. The cutting head was mounted in a small holder attached to the end opposite the needle. The head can pivot freely in this holder in a vertical direction only, and the whole arrangement can be held conveniently. When in position for testing, the weight of the unit is divided between the pivot (supported by the operator) and the needle resting on the subject, *see Figure 1*. If the unit is kept horizontal the applied weight is constant; it amounted to 80 grams. The unit can be moved rapidly from one spot to another with one hand, leaving the other free to operate the volume control and to make records of the readings obtained.

It is a common feature of apparatus of this kind that its response as amplitude of vibration at different frequencies is not uniform. That is, even if what is controlled and measured, in this case the voltage applied to the vibrator unit, is kept uniform, the response of the effective part, the vibrating needle, may not remain constant if the frequency of vibration is independently varied. This variation of response as a function of frequency alone may also be affected by the resistance and impedance of the skin and tissue to the movement imposed on it by the vibrating part; this tissue resistance may be different at different frequencies. In order to assess the size of any such variations at the different frequencies used the following was done. A small mirror was cemented to the needle of the vibrator. With the vibrator in place on a subject's toe, the mirror was illuminated so as to reflect a spot of light on a screen some distance

away. Then the movements of this spot of light were measured, when the needle was vibrating with constant voltage input to the vibrator, and the frequency of vibration was changed. The results as relative amplitude of movement, at the different frequencies, with constant voltage input to the vibrator, are given in Table 1.

It can be seen that from 150 C.P.S. to 600 C.P.S. the amplitude of the applied vibrations is relatively constant when the input voltage is kept the same. At the lower frequencies the amplitude increases, and above 600 C.P.S. the amplitude decreases.

The apparatus was well grounded at all times. For testing the toes it was arranged along the front of a table of normal height, and the subject was seated on the table to one side. The feet were supported on a low stand; part of the top of the stand facing the subject sloped down, the rest was flat. The feet were placed so that the toes more or less curled over the break between the sloping and horizontal portions.

The vibrator was used on the skin over the nail bed. In nearly every case the fifth or small toe was not tested. In a great many people this toe is so curved that it was not possible to place the vibrator on the toe in the proper position.

The subject was first allowed to feel the vibrations of one or two of the lower frequencies on a finger. He was instructed to keep his eyes closed throughout the test. With the volume control turned down, and no signal in the vibrator, the vibrator was placed on one of the toes. Then, at the frequency which had been chosen, the volume control was turned up until the subject reported that he felt vibration by saying "Yes." The strength at this threshold was noted. This was repeated with the other toes of the foot at the various frequencies, choosing toes at random, and frequencies irregularly. The usual procedure was to quickly and rather roughly map out the higher frequency limits, then determine the limits more carefully and fill in the intermediate frequencies.

At the beginning of the tests the subjects were warned that at

times they would feel no vibration, only the pressure of the instrument. Throughout the test blank stimuli were given—that is, the vibrator was placed on a toe for the usual interval but the volume control was not turned up. Most subjects of course made no response to this procedure. There were, however, a number of people who persisted in giving a positive response to such a blank stimulus and one is forced to conclude that either they were lacking in vibration sense under the conditions of the test or they so easily confused the touch or pressure of the instrument with the sensation of vibration that they could not distinguish between the two. All such cases were marked unreliable and were rejected. This method we believe to have been effective in ruling out false responses.

There are one or two points of a general nature about precautions which were taken. One of these concerns adaptation.

Cohen and Lindley (6) have studied adaptation to vibratory stimulation. They found that a supra-liminal stimulus of 60 C.P.S. raised the threshold. Stimulation at or near threshold strength had little effect upon the threshold unless continued for some time.

Evidence has been obtained that adaptation may be excluded under the conditions, and with the methods, used in this work.

The vibrator, clamped by its holder in a stand, was placed on the right great toe of each of three subjects and the threshold determined at a frequency of 200 C.P.S. by means of three consistent readings. The vibrator was left on at a little above threshold strength. At intervals, without moving the vibrator, the input was shut off and the threshold at the same spot quickly redetermined with three consistent readings. The vibrator was then turned on again, usually with an increase of strength to take care of the increase in threshold which was found.

In all three cases, the threshold rose somewhat during the first fifteen or twenty minutes and then remained constant. The total time was about thirty minutes, which is about as long as anyone cares to stand the application of the vibrator to one spot. In two

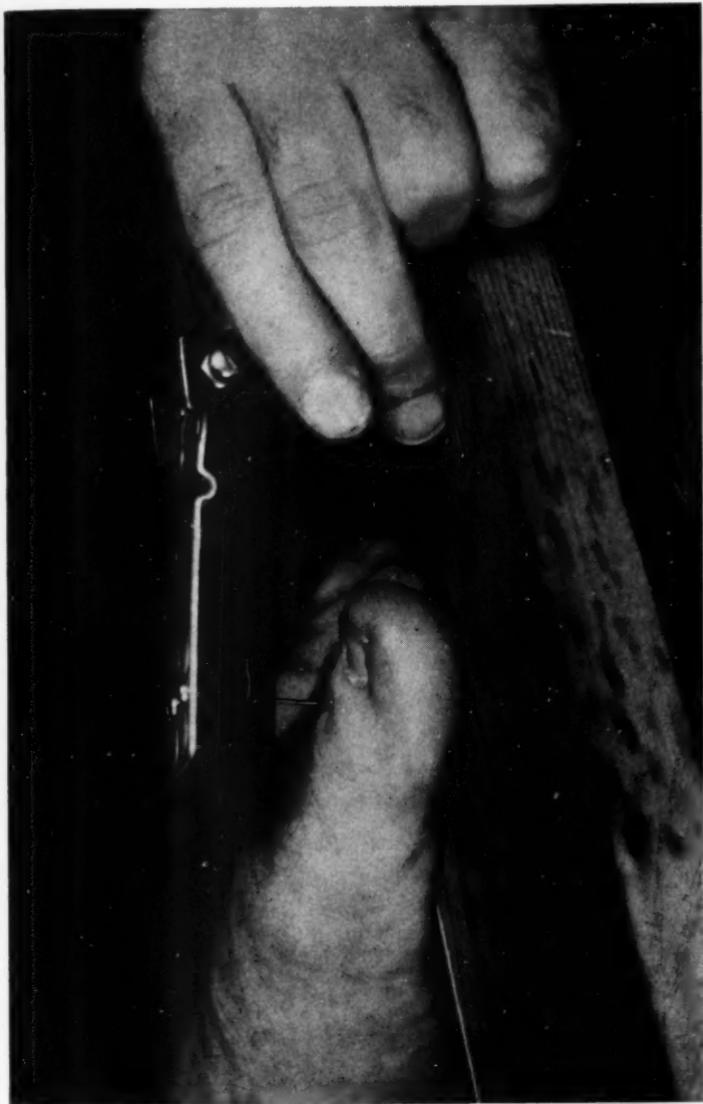


Fig. 1. Method of holding and applying vibrator unit.

M

ca
ca
ca
time
ad
mi
the
an

ing
as
tat
de
on
int
do
the
du
of

l
and
roo
tin
beg
irr
the
fee
rai
pre
me

tus
the
wi

cases the threshold was finally raised about 50 per cent and in one case about 100 per cent. At the end of the period of almost continuous stimulation, the vibrator was shut off. The threshold at the adapted spot returned to its original value in from four to eleven minutes. Although the threshold rose most in the first few minutes the rise in the first few seconds is too small to be determined with any certainty.

Thus, as continuous vibration, except for short intervals for testing the threshold, took several minutes to raise the threshold and as the return to normal was much quicker, it is unlikely that adaptation interfered in any way with the thresholds which have been determined. In the regular routine the vibrator was applied to any one toe for only a few seconds, after which there was a much longer interval while other toes were being tested. Moreover, the tests were done with the vibrator starting at rest, increasing the strength of the stimulation until the threshold was reached, so that the period during which vibration was appreciated was only a brief portion of the total time vibration was applied.

Many of the measurements were made during cold, rainy weather, and some on men who had been working outside. Although the room used was adequately heated the subject's bare feet were routinely warmed over a small electric heater for a short time before beginning. Feet which felt cold to the touch were found to give irregular, unreliable responses. It is very easy to show, by dipping the hand or foot into cold water and leaving it there until the skin feels somewhat cool to the touch, that this amount of chilling will raise the threshold considerably, see also Weitz (7). The need for precautions against chilling when vibration threshold measurements are to be taken cannot be over emphasized.

The group of over 400 men (1) who were tested with this apparatus were tested at the same time with tuning forks. Table 2 shows the relation between the tests with a tuning fork at 256 C.P.S. and with the electrically driven vibrator operated at 250 C.P.S. For the

	TOTAL TESTED	BOTH POSITIVE	BOTH NEGATIVE	VIBRATOR NEG. AND TUNING FORK POS.	VIBRATOR POS. AND TUNING FORK NEG.
Number of Toes	1,692	1,323	47	10	312
Per Cent of Toes	100	78.2	2.8	0.6	18.4

Table 2. Comparison of positive and negative responses to tuning fork at 256 C.P.S. and to electric vibrator at 250 C.P.S. applied to four toes of each man tested.

purposes of this comparison the results from the electric vibrator have been used only to indicate whether the subject being tested did or did not respond. A response at any amplitude was considered positive. A negative response meant that he made no response to even the maximum stimulus which the instrument could give. The tests with the tuning fork and the electric vibrator were given by different persons. As the men were tested on four toes, rather than taking any arbitrary figure of, say, three toes positive, and calling such a case positive, or adopting any other measure of dividing positive from negative cases, tabulation has been made of the whole number of toes tested, which gives a figure four times the number of cases. As can be seen the agreement between the two tests is good, and in 81 per cent of the cases the results were the same. There is a small number of cases positive to 256 C.P.S. tuning fork and negative to 250 C.P.S. vibrator, and a larger number negative to 256 C.P.S. fork and positive to 250 C.P.S. vibrator. It should be noted that the electric vibrator, working at its full strength, is capable of giving a somewhat stronger stimulus than the tuning fork, so that this last result is to be expected.

Besides this positive or negative type of response, the thresholds at the various frequencies were found on over 200 men. Of the 400 men included in Table 2 some were excluded from this test because their response at some frequency or other was considered unreliable by the criterion given above for false responses. Of the remainder, the thresholds were obtained for the most sensitive toe of the least sensitive foot.

The thresholds are given in Table 3 in numbers ranging from 0 to 7, derived from the voltmeter on the instrument in the way described. All the readings so obtained at each of the frequencies used have been averaged. Values in Table 3 are mean thresholds. The results have been divided into three classes according to the age of the men. The table also gives the standard errors of the means, the number of men tested at each frequency, and the statistical significance of the differences between the means of the age groups is indicated by the ratio of a difference to the standard error of the difference.

The thresholds of each age group present a fairly regular varia-

Table 3. Mean vibration thresholds at various frequencies for employed men.
Comparison by ages.

AGE GROUP IN YEARS	FREQUENCY, CYCLES PER SECOND									
	50	100	200	250	300	350	400	500	600	800
I 20-29										
Mean Thresholds	3.7	2.4	2.9	3.6	4.2	5.4	5.0	5.6	5.9	6.2
St. Error of Mean	.04	.06	.07	.06	.06	.04	.06	.05	.06	.07
Number Tested	276	277	276	273	273	90	267	237	177	82
II 30-39										
Mean Thresholds	4.0	2.8	3.4	4.1	4.6	5.5	5.3	5.8	6.1	6.3
St. Error of Mean	.08	.09	.11	.09	.10	.09	.09	.09	.09	.12
Number Tested	120	116	126	125	121	56	112	90	65	31
III 40+										
Mean Thresholds	4.4	3.4	4.0	4.7	5.1	5.7	5.6	5.8		
St. Error of Mean	.17	.24	.25	.28	.17	.18	.22	.22		
Number Tested	25	27	26	26	24	13	18	11		
Ratio of Difference to St. Error of Diff.										
I and II	2.91	3.18	4.13	4.76	3.83	1.49	2.78	2.04	1.55	.29
II and III	2.33	2.59	1.94	3.08	2.42	0.75	1.24			
I and III	3.90	4.12	4.08	5.87	5.00	1.58	2.58	1.14		

As stated in the text some men have been excluded because their responses were inconsistent and, therefore, unreliable. These men do not appear in the table. Variations in the numbers of men tested at different frequencies, other than tapering off at high frequencies caused by the approach of the threshold curve to the maximum stimulus curve, merely show that not every man was tested at all the frequencies listed. The upper and lower frequency limits of each man were always found, intermediate frequencies were not always completely covered.

tion with frequency. They are higher at the lowest and highest frequencies than in the neighborhood of 100-200 C.P.S. It should be pointed out that the characteristics of the vibrator unit are such (see Table 1), that the thresholds at 50 and 100 C.P.S. are higher than the figures show. The reason for the drop at 400 C.P.S. is undetermined.

This variation of threshold with frequency is similar to previously reported results, but in general is based on a larger number of cases. The shape of this curve has a bearing on the results at the upper frequencies. The output of the vibrator has an upper limit of strength, and as one extends the tests to the higher frequencies a point is reached where this curve of declining maximum strength is intersected by the rising curve of the threshold to sensation. At frequencies beyond this point there will be no response, though the subject might still be aware of a sensation if a stronger stimulus could be used.

The table shows that there is a rise in the threshold with increasing age. This is one of the most striking features, and was one of the first effects to be noticed. Newman and Corbin (8) seem to have been the first to measure this quantitatively, though it had been observed previously (9). They measured the threshold at one frequency, 60 C.P.S. The present work, based on a large number of men, shows increase of threshold with age in all three age groups, and at all frequencies used. The effect of age is to raise the whole threshold curve so that the point at which it intersects the curve of maximum signal strength falls at a lower frequency. It can be seen that the older men fail to respond to the higher frequencies. At the highest frequencies in each age group the means all tend to approach the same upper limit; so the differences are small. At 350 C.P.S. fewer men were tested. With these two exceptions nearly all the differences between mean thresholds of the different age groups are significant, that is the ratios of the differences of the means to the standard error of the differences are more than 2.

With this method a subject's vibration thresholds may be found readily over the whole range of frequencies. All that is necessary is to set the frequency control to the desired frequency, hold the vibrator unit in place, turn up the volume control till the subject responds and read the voltmeter at this point. This gives a number between 0 and 7, which is related to the amplitude as described above, and this number is the threshold. Some practice with normal young people whose responses are quick and definite is helpful. In older people the range of effective frequencies is generally limited and more care is needed to exclude false responses to stimuli outside this range, some of which must be given to establish the limits. Even within the effective range their responses are apt to be hesitant. Explanation at the start, together with several applications of frequencies in the 100-200 C.P.S. range, of sufficient strength to produce adequate stimulation, should be used.

It is not possible to state whether the ability to measure thresholds at several frequencies will prove to be of practical significance. The present evidence would suggest that determinations at two or three frequencies, corresponding to the frequencies of the commonly used tuning forks, would be sufficient, and would differentiate between the different vibratory sensitivities of individuals as well as a more extended series. All the men examined were employed and healthy enough to be at work. In cases of disease, in which changes in vibratory sensitivity might be expected, and particularly those in which subsequent recovery might follow, it might be of value to follow the thresholds at several frequencies. The shape of the curve relating the thresholds of vibration to frequency, is similar to the curve for the variation of the threshold of hearing with frequency, and in the latter case losses of hearing are not always uniform throughout the audible frequency range.

SUMMARY

An electrically driven apparatus is described which is suitable

for measuring the responses in man to mechanical vibrations of different frequencies.

The apparatus was used in testing several hundred men, and the technique is described.

Results are given which show the variation of the thresholds with respect to the frequency of vibration, and their increase with increasing age.

REFERENCES

1. Nutritional Status of Aircraft Workers in Southern California. IV. To be published in *The Milbank Memorial Fund Quarterly*.
2. Yaczynski, G. K. and Brown, Meyer: Studies of the Sensation of Vibration; I. Variability of the Vibratory Threshold as a Function of Amplitude and Frequency of Mechanical Vibration. *Journal of Experimental Psychology*, 1941, 28, pp. 509.
3. Fox, James C. Jr. and Klemperer, Wolfgang W.: Vibratory Sensibility, A Quantitative Study of Its Thresholds in Nervous Disorders. *Archives of Neurology and Psychiatry*, 1942, 48, pp. 622.
4. Terman, F. E.; Buss, R. R.; Hewlett, W. R.; and Cahill, F. C.: Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment. *Proceedings of the I.R.E.*, 1939, 27, pp. 649.
5. THE "RADIO" HANDBOOK, 1941, 8th. edition, pp. 514.
6. Cohen, Louis H. and Lindley, Stanley B.: Studies in Vibratory Sensibility. *American Journal of Psychology*, 1938, 51, pp. 44.
7. Weitz, Joseph: Vibratory Sensitivity as a Function of Skin Temperature. *Journal of Experimental Psychology*, 1941, 28, pp. 21.
8. Newman, Henry W. and Corbin, Kendall B.: Quantitative Determination of Vibratory Sensibility. *Proceedings of the Society for Experimental Biology and Medicine*, 1936, 35, pp. 273.
9. Pearson, Gerald H. J.: Effect of Age on Vibratory Sensibility. *Archives of Neurology and Psychiatry*, 1928, 20, pp. 482.

SOCIAL AND PSYCHOLOGICAL FACTORS AFFECTING FERTILITY

V. THE SAMPLING PLAN, SELECTION, AND THE REPRESENTATIVENESS OF COUPLES IN THE INFLATED SAMPLE¹

P. K. WHELPTON AND CLYDE V. KISER

A. THE SAMPLING PLAN

AS pointed out in the preceding article of this series, a major purpose of the Study is to determine and evaluate the reasons why fecund married couples with zero, one, two, or other number of live births have this number rather than more or fewer, especially if the size of the family is planned. For such research it is important that a sufficiently large number of schedules be completed for fecund couples with each number of live births to meet the requirements of statistical analysis. Since budgetary considerations limited the total number of fecund couples for whom schedules could be filled out, the problem was to distribute this total most effectively by size of family. The information which was available when plans for the field work were being prepared showed that among the couples meeting the demographic and educational requirements for inclusion in the Study² there probably would be (a) a somewhat larger number with one or two live births than with none, and (b) a substantially smaller number with three live births than two, with four than three, etc. Other information indicated that the proportion of families planned as to size probably would vary in the same direction, but more abruptly. It appeared desirable, therefore, to sample the eligible couples with zero, one, or two live births, and to complete schedules for as many as possible of those with three or more.

¹This is the fifth of a series of reports on a study conducted by the Committee on Social and Psychological Factors Affecting Fertility, sponsored by the Milbank Memorial Fund with grants from the Carnegie Corporation of New York. The Committee consists of Lowell J. Reed, Chairman; Daniel Katz; E. Lowell Kelly; Clyde V. Kiser; Frank Lorimer; Frank W. Notestein; Frederick Osborn; S. A. Switzer; Warren S. Thompson; and P. K. Whelpton.

²For a list of these requirements, see *infra*, pp. 86 and 87.

According to the schedules which were filled out for the white occupants of 102,499 dwelling units in the Household Survey of Indianapolis, there were 2,589 couples meeting the requirements for inclusion in the detailed Study.⁸ Of these, 529 reported no live birth, 727 reported one, 801 two, 310 three, and 221 four or more live births.⁹ If the Study had been equally interested in all couples in each size of family group these data would have sufficed for establishing the sampling ratios. In order to take into account the effects of sterility and the planning of family size, however, it was necessary to await the availability of data gathered during the first months of the Study.

On September 20, 1941, when approximately half of the field work was finished,¹⁰ the couples for whom schedules had been completed were subdivided into three groups, namely, (1) relatively fecund,¹¹ size of family planned; (2) relatively fecund, size of family quasi-planned or too large;¹² and (3) relatively sterile. Relatively sterile couples included all with three or fewer live births who knew, or had good reasons for believing, that during a consecutive period of two or three years since marriage it was physiologically impossible for them to have a child.¹³ Relatively fecund couples in-

⁸ These figures do not include the 339 occupied dwelling units nor the seventeen eligible couples in Tract 103. This tract has a northern boundary almost one mile south of the remainder of Indianapolis (to which it is connected by only one street within the City limits) and is an important market-gardening area. For various reasons it was excluded from the detailed Study.

The foregoing and most of the subsequent data regarding the Study are from machine tabulations of punch cards, and in some instances differ slightly from the results of hand counts made during the field work.

⁹ The number of live births to one couple was not reported in the Household Survey.

¹⁰ The field work began on April 15, 1941, and ended on January 31, 1942.

¹¹ According to definitions adopted by the Population Association of America, fecundity means the *physiological ability* to participate in reproduction as distinguished from birth performance; sterility means the absence of such ability. These terms (modified by "relatively," as explained below) are used in this paper in accordance with these definitions.

¹² Size of family was considered "too large" if the wife and/or the husband did not want the last pregnancy ending in a live birth either when it occurred or later.

¹³ Failure to conceive in the absence of regular contraceptive practice during two or three consecutive years (two for couples with no pregnancy and three for others) was the chief criterion in establishing "good reason for believing" that conception was physiologically impossible. Since this criterion relates to any given period of two or three consecutive years, it is not surprising that a large proportion of the couples classified as relatively sterile actually had children.

cluded all couples with four or more live births, and those with three or fewer who were not classified as relatively sterile.⁹ Relatively fecund couples were considered to have planned family size (a) if no conception had occurred (presumably because of the regular practice of contraception); (b) if contraceptive practices had been discontinued a month or more prior to the interview in order that conception might occur; or (c) if the last conception occurred when contraception was stopped for that purpose, or when it was practiced "sometimes" after having been practiced "usually" or "always" and both the wife and husband said that they wanted a child at that time.¹⁰ Other relatively fecund couples were classified as having size of family quasi-planned or too large. The results of the September 20th classification are shown in the upper half of Table 1. They indicated that all couples with three or more live births should be interviewed, and that a 75 per cent sample of childless couples and a 50 per cent sample of couples with one or two children should yield approximately equal numbers of relatively fecund couples planned as to family size. (See Table 1, line 15.) These sampling ratios were adopted at that time, and provided what will be referred to hereafter as Sample A.

Because of the important relation between socio-economic status and fertility, it was desirable to keep the sample within each size of family group similar to the entire group of eligible couples of the same parity with regard to socio-economic status. Of the three criteria available from the Survey schedules — highest grade of

⁹ Couples with four or more live births "who knew or had good reason for believing that during a consecutive period of three years since marriage it was physiologically impossible for them to have a child" were included in the "relatively fecund" group because (a) the number of couples with four or more live births in the total eligible group was small, and (b) the fertility (i.e., birth performance) of these "relatively sterile" couples was higher than that of the large majority of "relatively fecund" couples.

¹⁰ In this connection a plain water douche immediately after intercourse was not considered a contraceptive practice for a wife who insisted that it was "for cleanliness only," or (referring to a cold water douche) that it was "to help me get pregnant."

The basis of classifying couples as to planning size of family which is being used in the analysis differs slightly from that used in hand counts while the interviewing was progressing, and will be explained in a later article.

school completed by wife and husband, tenure, and rent paid for a rented home or estimated rental value of an owned home — the latter seemed most suitable. The process was that of applying the percentage rental distribution of all eligible families of each size to the number of families of corresponding size desired for the sample. Subtracting the number of couples already interviewed from the number desired in each fertility-rental category gave the number to be interviewed after September 20. (See Table 2.) Specific couples were chosen for interview by arranging the Survey schedules for

Table 1. Couples seen or "lost" before September 20, 1941, by number of live births, completion of schedules, fecundity and planning of size of family; sampling ratios for Sample A and couples in Sample A¹ (classified as above).

STUDY STATUS	No LIVE BIRTH		ONE LIVE BIRTH		TWO LIVE BIRTHS		THREE OR MORE LIVE BIRTHS	
	Num- ber	Per Cent	Num- ber	Per Cent	Num- ber	Per Cent	Num- ber	Per Cent
1 Seen or "Lost" Before September 20 ²	120	100.0	248	100.0	264	100.0	207	100.0
2 Schedules Not Completed ³	37	30.8	80	32.3	74	28.1	57	27.6
3 Schedules Completed	83	69.2	168	67.7	190	71.9	150	72.4
4 Relatively Sterile	53	44.2	51	20.6	37	14.0	16	7.7
5 Relatively Fecund	30	25.0	117	47.1	153	57.9	134	64.7
6 Size of Family Planned	30	25.0	77	31.0	74	28.0	29	14.0
7 Size Quasi-Planned, or Too Large	0	—	40	16.1	79	29.9	105	50.7
8 Eligible According to Survey ⁴	529	—	727	—	802	—	532	—
9 Sampling Ratios For Sample A	—	75.0	—	50.0	—	50.0	—	100.0
10 Couples in Sample A ⁵	307	100.0	364	100.0	400	100.0	531	100.0
11 Schedules Not To Be Completed ⁶	122	30.8	118	32.3	112	28.1	147	27.6
12 Schedules To Be Completed ⁷	275	69.2	246	67.7	288	71.9	384	72.4
13 Relatively Sterile ¹	176	44.2	75	20.6	56	14.0	41	7.7
14 Relatively Fecund	99	25.0	171	47.1	232	57.9	343	64.7
15 Size of Family Planned	99	25.0	113	31.0	112	28.0	74	14.0
16 Size Quasi-Planned, or Too Large	0	—	58	16.1	120	29.9	269	50.7

¹ Sample A is based on the sampling ratios on line 9 of this table. In Sample A the schedules of the "deferred" couples (i.e., 3 out of 4 couples with no live birth, classified as relatively sterile, interviewed after September 20, 1941, and apparently willing to cooperate, but for whom the interviewing was terminated with Form A for reasons explained in the text) are considered "completed."

² A couple is considered "seen" if the interviewer spoke to some member of the household, and "lost" if she learned from neighbors or others that the family had moved out of Indianapolis, or to an unknown address.

³ Consists primarily of couples who were found to be ineligible, who refused to cooperate, or who had moved out of Indianapolis or to an unknown address since the Survey. See Section B of this article.

⁴ Excludes one couple for whom the number of live births was not reported in the Household Survey.

⁵ In each parity the percentage distribution of the couples "seen or lost" before September 20¹ is used in subdividing the "Couples in Sample."

Table 2. All couples, and couples in Sample A¹, by number of live births and monthly rental value of dwelling unit.

RENTAL VALUE OF DWELLING UNIT	ALL COUPLES		COUPLES IN SAMPLE A		
	Number	Per Cent	Number ²	Interviewed Before Sept. 20	To Be Interviewed After Sept. 20 ³
NO LIVE BIRTH					
TOTAL	529	100.0	397	120	277
Under \$10	47	9.1	36	15	21
\$10-24	34	6.6	26	10	16
25-29	60	11.6	46	15	31
30-34	73	14.1	56	18	38
35-39	87	16.8	67	24	43
40-49	86	16.6	66	17	49
50-69	94	18.1	72	14	58
70-89	9	1.7	7	0	7
90 or More	10	1.9	7	1	6
Relatives ⁴	18	3.5	14	6	8
Lodgers ⁵	2	—	—	—	—
No Data ⁶	9	—	—	—	—
ONE LIVE BIRTH					
TOTAL	727	100.0	364	248	116
Under \$10	81	11.2	41	32	9
\$10-24	60	8.3	30	21	9
25-29	111	15.4	56	42	14
30-34	102	14.1	51	35	16
35-39	103	14.3	52	46	6
40-49	98	13.6	50	38	12
50-69	108	15.0	55	26	29
70-89	35	4.8	17	1	16
90 or More	12	1.7	6	1	5
Relatives ⁴	12	1.7	6	6	0
No Data ⁶	5	—	—	—	—
TWO LIVE BIRTHS					
TOTAL	801	100.0	400	264	136
Under \$10	119	14.9	60	36	24
\$10-24	86	10.8	43	32	11
25-29	112	14.0	56	48	8
30-34	106	13.3	53	34	19
35-39	117	14.6	59	34	25
40-49	92	11.5	46	25	21
50-69	98	12.3	49	29	20
70-89	46	5.8	23	13	10
90 or More	14	1.7	7	6	1
Relatives ⁴	9	1.1	4	7	-3
No Data ⁶	2	—	—	—	—

¹ See Table 1, footnote 1.

² The "Total" lines are from Table 1, line 10. The other lines are computed by multiplying the totals by the percentages in the column to the left.

³ The difference between the two columns to the left.

⁴ Living with relatives, monthly rental of dwelling unit not stated.

⁵ Omitted from computation of per cents and from Sample A.

couples not yet interviewed by tract and block number within each fertility-rental category, and selecting in rotation.

Among couples with no live birth, those classified as relatively sterile were found to outnumber by nearly two to one those classified as relatively fecund, size of family planned. (*See Table 1.*) In consequence, the sampling ratio needed for the fecund childless couples was much larger than that needed for the sterile. Since the Study was directed primarily at relatively fecund couples, it was decided to "defer" three-fourths of the sterile couples with no live birth. Accordingly, each interviewer was instructed (a) to keep a list of couples with no live birth who appeared to have been sterile for two or more years judging from the replies to two questions on Form A (the brief introductory form), and (b) to fill out Form S (for sterile couples) for only the last of each four of these couples. Modifying Sample A in this manner yielded Sample A-1.¹¹

The extent to which the couples who were interviewed up to September 20 are typical of those in Sample A depends in large part on how the interviewers had been assigned to various areas of the City. From April 15 (when field work began) to July 1 an attempt was made to work in census tracts of all types. During July and August, however, attention was concentrated on the middle and lower economic areas because of the direct relation which was believed to exist between economic status and summer vacations. As a result, in each parity the ratio of couples to be interviewed after September 20 to those already interviewed varied directly with monthly rental. (*See Table 2.*) For example, among couples with one live birth only about 25 to 30 per cent of the number desired in the rental groups under \$30 remained to be interviewed, as compared with over 50 per cent of the number desired in each of the rental groups above \$50. Because it was believed that a direct rela-

¹¹ In referring to samples the capital letter A or B denotes which sampling ratios were used in assigning couples to interviewers. The figure 1 after the letter denotes exclusion from the sample of the "deferred" sterile couples for whom only a short schedule (Form A) was filled out; absence of the figure 1 denotes the inclusion of such couples.

tion exists between economic status and the planning of size of family, it was expected that the larger proportion of couples in the upper economic groups among those to be interviewed after September 20 would tend to raise somewhat the percentage of relatively fecund couples planned as to size of family, shown in Table 1, line 15.

When about 90 per cent of the field work was finished (on January 4, 1942) a second classification was made of the couples for whom schedules had been completed to date. It showed that the proportion of such couples among all couples in Sample A was substantially lower for those with no live birth than for those with one or more.²² Although the interviewers had tried more diligently to see couples with no live birth than those with one or more, they had been less successful because of the greater difficulty of finding the wife or husband at home. As a result, the number of schedules completed for "relatively fecund, size of family planned" couples with no live birth was well below the corresponding number for similar couples with two live births. (*See Table 3, line 11.*) The number of schedules completed for "relatively fecund, size of family planned" couples with three or more live births was even smaller, primarily because the proportion of such couples who planned family size was low as compared with proportions for couples having fewer than three live births. To improve the situation the interviewers were instructed to give priority thenceforth to couples with no live birth and those with three or more. To make possible their completing the additional schedules desired for the former, the sampling ratio for childless couples was expanded from 75 per cent to 100 per cent. The sample resulting from the January 4th modifications will be referred to as Sample B or B-1.²³ It was realized

²² Schedules had been completed for 52 per cent of the couples with no live birth in Sample A (206 of 397) and for 66 per cent of the couples with one or more live births in Sample A (856 of 1,295). (*See Table 3, lines 1 and 7.*) In this connection, the schedules of the "deferred" sterile couples are considered as having been "completed" although only Form A was filled out.

²³ *See* footnote 11.

that this change in sampling procedure would result in schedules being filled out for a larger proportion of the childless couples easy to find at home than of other childless couples, and that differences in this characteristic probably are related to differences in certain others, such as employment of wife. It was believed, however, that the biases which might be introduced by raising the sampling ratio for childless couples to 100 per cent would be a lesser evil than obtaining schedules from too few fecund childless couples.

The January 4th classification showed also that for each size of family group the distribution by rent or rental value for couples for whom schedules were completed differed in certain respects from that for all couples in Sample A. (See Table 4.) Among couples with no live birth the proportions in the \$20-24 and \$50-69 rental classes were somewhat lower for those for whom schedules had been completed than for others, but the proportions in the \$30-39 classes were somewhat higher for the former. Such differences resulted

Table 3. Couples seen or "lost" before January 4, 1942, by number of live births, completion of schedules, fecundity, and planning of size of family.

STUDY STATUS	No LIVE BIRTH		ONE LIVE BIRTH		TWO LIVE BIRTHS		THREE OR MORE LIVE BIRTHS	
	Num- ber	Per Cent	Num- ber	Per Cent	Num- ber	Per Cent	Num- ber	Per Cent
1 In Sample A ^{1,2}	397	100.0	364	100.0	400	100.0	531	100.0
2 Not Seen or "Lost" Before Jan. 4, 1942 ³	103	25.9	24	6.6	43	10.8	80	15.1
3 Seen or "Lost" Before Jan. 4, 1942 ⁴	294	74.1	340	93.4	357	89.2	451	84.9
4 Seen or "Lost"	294	100.0	340	100.0	357	100.0	451	100.0
5 Schedules Not To Be Completed ⁴	88	29.9	95	27.9	84	23.5	113	25.1
6 Schedules Completed	206	70.1	245	72.1	273	76.5	338	74.9
7 Schedules Completed	206	100.0	245	100.0	273	100.0	338	100.0
8 Relatively Sterile ⁵	132	64.1	71	29.0	44	16.1	20	5.9
9 Relatively Fecund	74	35.9	174	71.0	229	83.9	318	94.1
10 Relatively Fecund	74	100.0	174	100.0	229	100.0	318	100.0
11 Size of Family Planned	74	100.0	82	47.1	97	42.4	59	18.6
12 Size of Family Quasi-Planned or Too Large	0	0.0	92	52.9	132	57.6	259	81.4

¹ See Table 1, footnote 1.

² From Table 1, line 10.

³ See Table 1, footnote 2.

⁴ See Table 1, footnote 3.

RENTAL VALUE OF DWELLING UNIT	NO LIVE BIRTH		ONE LIVE BIRTH		TWO LIVE BIRTHS	
	Couples in Sample A ¹	Schedules Completed	Couples in Sample A ²	Schedules Completed	Couples in Sample A ³	Schedules Completed
Number of Couples	397	206	364	245	400	273
<i>Per Cent Distribution:</i>						
Total	100.0	100.0	100.1	100.0	100.0	100.0
Under \$20	9.1	10.0	11.2	9.8	14.9	15.8
\$20-24	6.6	4.5	8.3	8.6	10.8	9.6
25-29	11.6	11.9	15.4	15.5	14.0	15.1
30-34	14.1	15.9	14.1	17.1	13.3	14.7
35-39	16.8	18.9	14.3	15.1	14.6	16.9
40-49	16.6	16.9	13.6	13.1	11.5	12.1
50-69	18.1	15.4	15.0	12.7	12.3	9.9
70-89	1.7	2.0	4.8	5.7	5.8	4.8
90 or More	1.9	2.0	1.7	.8	1.7	.4
Living With Relatives	3.5	2.5	1.7	1.6	1.1	.7

¹ See Table 1, footnote 1.² From Table 2.

Table 4. Couples in Sample A¹ and couples for whom schedules were completed before January 4, 1942, by number of live births, and monthly rental value of dwelling unit.

from variations between rental groups in the ease of finding couples at home, in the willingness of couples to cooperate, and in other factors.²⁴ In order to keep the rental distribution of cooperating childless couples as similar as possible to that of all childless couples, the interviewers were asked to try to see couples in the relatively under-represented rental groups and only as a last resort to call on those in the relatively over-represented groups. Among couples with one live birth a similar situation was met by giving Sample B thirty-two more couples than Sample A, and selecting them from the rental groups whose percentages needed raising. Since the number of schedules completed for "relatively fecund, size of family planned" couples was largest for those with two live births, Sample B was given only five more two-child couples than Sample A, and improvement of the rental distribution was sought primarily through withdrawing a small number of two-child couples in the rental groups whose percentages needed lowering and replacing

²⁴ These variations will be discussed in Section B.

them by an equal number of couples in the rental groups needing additional representation.²⁸

Although the interviewers were unable by the date set for terminating field work to complete schedules for all of the couples added to Sample A on the basis of the January 4 inventory, progress toward the goal just outlined was achieved. The number of "relatively fecund, size of family planned" couples for whom schedules were completed was raised to eighty-eight for those with no live birth and to sixty-four for those with three or more, as compared with eighty-three and 102, respectively, for those with one and two.²⁹ Furthermore (as will be brought out in Section D), in each size of family group the differences between (a) the rental distribution of the couples for whom schedules were completed, and (b) that of all eligible couples in the Survey, were reduced between January 4 and the termination of the field work.

B. THE CAUSES, EXTENT, AND EFFECTS OF SELECTION

The preceding section refers to sampling, which was voluntary selection performed in accordance with a plan prepared by the Committee and field staff. This section refers primarily to involuntary selection, which was not desired but which could not be prevented. The most important causes of this selection are (a) some couples were "lost" to the Study because they moved out of Indianapolis or to an unknown address; (b) there was no way of locating couples who moved to Indianapolis after the Survey but before the Study interviewing ended; (c) the interviewers did not try to see

²⁸ Four other substitutions in Sample A (or A-1) had been made previously because (a) one of the interviewers resigned and failed to complete schedules which she had begun for four couples, and (b) it was not believed wise to change interviewers when the schedules were partially completed. These substitutions were made within the rental group and (if possible) within the tract.

²⁹ The foregoing is a distribution of fecund planned families by number of live births reported in the Household Survey. The distribution based on the Study schedules is as follows: eighty-eight fecund planned couples with no live birth, sixty-six with three or more, seventy-seven with one, and 106 with two. The slight differences between the two distributions are due to (a) errors in reporting number of live births in the Survey, and (b) the classification of adopted children as "live births" to the couple in the Study but not in the Survey.

some couples because they were busy with others, or they tried but found no one at home; (d) there were errors in the information secured in the Survey and used as a basis for determining eligibility; and (e) there were refusals to cooperate in the Study.

Of the 1,865 couples who were supposed to be interviewed in accordance with Sample B discussed in the preceding section,²⁷ thirty-eight were "lost" because they moved away from Indianapolis and fifty because they moved to an unknown address. In most cases the move occurred before a call was made by an interviewer; in a few cases it was after the first interview (at which only Form A was filled out) but before the second could be arranged. Since these eighty-eight couples (as a group) have a higher rate of moving than the remaining 1,777 couples in Sample B, they presumably differ in other respects. As far as the data on the Household Survey schedules are concerned, however, the only significant differences between the two groups relate to tenure and State of birth. As would be expected, the proportion of couples owning their home is much lower for the "lost" couples (18.2 per cent) than for the other couples (41.8 per cent). (See Table 5.) Similarly, the proportion of wives and husbands who were born in Indiana is significantly lower for the "lost" couples (62.5 and 58.0 per cent) than for the others (71.7 and 70.2 per cent). That there are few if any significant differences between the two groups in attitudes toward family size and its control is indicated by the similarity in average number of live births (1.6 and 1.7). It seems reasonable to conclude, therefore, that the Study was not biased seriously by the failure to complete schedules for the thirty-eight couples who moved out of Indianapolis or for the fifty who moved to an unknown address. Because of the net migration to Indianapolis which is believed to have occurred during 1941, it is probable that between forty and seventy-five couples meeting the eligibility requirements of the Study moved to Indianapolis after their neighborhoods had been canvassed by

²⁷ The 1,865 couples include the "deferred" sterile couples.

the Survey but before the date of termination of the field work for the Study. It seems probable that as a group they tended to resemble the thirty-eight couples who are known to have moved out of Indianapolis during this period, and that excluding them from the group interviewed did not introduce any serious biases.

In addition to the eighty-eight couples who were not seen because they moved, 129 were not seen either because the interviewers could not find anyone at home, or because they were busy with other couples and did not call. Two-thirds (eighty-three of 129) of the couples not found at home or not called on were childless. An important proportion of them were added to Sample A (or A-1) on January 4 and lived in the areas of the City where the interviewers had worked previously. These couples were not seen because the

Table 5. Characteristics of (a) couples in Sample B but not seen because they moved out of Indianapolis or to an unknown address, and (b) other couples in Sample B.³

CHARACTERISTICS	COUPLES NOT SEEN BECAUSE THEY MOVED	OTHER COUPLES IN SAMPLE B
Number of Couples	88 ²	1,777 ²
Average:		
Number of Live Births	1.6	1.7
Age { Wife	33.9	34.2
Husband	36.7	37.0
Date of Marriage ⁴	8-9-28	7-23-28
Highest Grade of School		
Completed ⁵ { Wife	11.2	11.2
Husband	11.4	11.1
Rental Value of Dwelling Unit	\$36.97	\$35.57
Per Cent Owning Home	18.1	41.8
Per Cent Born in Indiana { Wife	62.5	71.7
Husband	58.0	70.1

¹ Sample B was obtained from Sample A by raising the sampling ratio for zero parity from 75 to 100 per cent and making certain additions and substitutions of couples with one or two live births (see text). It includes the "deferred" sterile couples.

² Monthly rent or rental value was not reported for one of these couples in the Survey.

³ Monthly rent or rental value was not reported for 53 of these couples in the Survey. For other items the number of "unknowns" varies between one and seven.

⁴ Because month and day of marriage were not asked in the Survey it is assumed in averaging year of marriage that all marriages occurred on July 2, the mid day of the year. In consequence the averages have a high margin of error.

⁵ High school is considered as consisting of grades 9 through 12, and college of grades 13 through 16.

interviewers spent most of their time after January 4 in other areas, where they could fill out the additional schedules desired for couples with three or more children as well as those with none.

Since the childless couples originally chosen for Sample A constituted a 75 per cent sample (stratified by rent) of all childless couples, the selection just described would not be expected to bias the results significantly. In contrast, certain biases may have been introduced by not "camping on the door step" of couples whom it was difficult to find at home. In some cases, of course, many calls were made. In others, the first call occurred when the interviewer had nearly completed her work in that area; hence subsequent calls soon became too time-consuming and were discontinued. It is probable that couples hard to find at home differ from others with respect to several of the conditions believed to affect fertility and therefore under investigation, for example, employment of the wife and interest in social activities. Unfortunately, however, the only differences which can be measured are those relating to the few items on the Survey schedules. Even here the couples not called on and those called on but not found at home must be considered together¹⁸ and the comparison within a parity must be confined to childless couples because of the small number of couples in the other groups.

An analysis of the data on the Survey schedules shows only one difference of any importance between the characteristics of childless couples who were not seen for reasons other than moving, and those of other childless couples. The percentage of homes owned is 32.9 for the former group and 40.1 for the latter (*See Table 6*), but the difference of 7.2 is too small to be statistically significant for the number of couples involved. If couples of all parities are combined, the differences between the two groups are relatively large for average number of live births and are significant statistically for

¹⁸ No distinction was made in coding, in part because of the small number of couples involved.

rent or rental value of home, but not for other items. To understand the meaning of these differences it must be remembered that (a) because of the expansion of the sampling ratio for childless couples from 75 to 100 per cent on January 4, the proportion of childless couples is much higher among couples not contacted for reasons other than moving (64.3 per cent) than among other couples in Sample B (25.7 per cent), and (b) there is an important indirect relation between the number of live births to a couple and the rent or rental value of their home; the rental value averages \$39.76 for

Table 6. Characteristics of (a) couples in Sample B but not seen because of reasons other than moving, and (b) other couples in Sample B.¹

CHARACTERISTICS	COUPLES WITH NO LIVE BIRTH		ALL COUPLES	
	Not Seen Because of Reasons Other Than Moving	Other Couples in Sample B	Not Seen Because of Reasons Other Than Moving	Other Couples in Sample B
Number of Couples	83 ²	446 ³	129 ⁴	1,736 ⁵
Average:				
Number of Live Births	—	—	0.6	1.8
Age { Wife	35.6	35.6	36.0	34.1
{ Husband	38.1	38.6	37.5	37.0
Date of Marriage ⁶	9-17-28	7-30-28	8-12-28	7-23-28
Highest Grade of School				
Completed? { Wife	11.8	11.2	11.6	11.1
{ Husband	11.8	11.3	11.6	11.1
Rental Value of Dwelling Unit	\$38.59	\$39.73	\$39.67	\$35.34
Per Cent Owning Home	32.9	40.1	37.0	41.0
Per Cent Born in Indiana { Wife	68.7	68.8	70.3	71.4
{ Husband	69.9	68.1	68.8	69.7

¹ See Table 5, footnote 1.

² Tenure was not reported for one of these couples in the Survey, and monthly rent or rental value for seven.

³ The number of "unknowns" in the Survey is 22 for monthly rent or rental value, and zero, one, or two for the other items.

⁴ The number of "unknowns" is seven for monthly rent or rental value, and zero or one for the other items.

⁵ The number of "unknowns" is 47 for monthly rent or rental value, and zero to seven for the other items.

⁶ See Table 5, footnote 4.

⁷ See Table 5, footnote 5.

childless couples eligible for the Study and \$35.95 for couples with children. In other words, when all parities are combined, the lower number of live births and the more expensive housing of the couples not seen for reasons other than moving certainly is due chiefly, and perhaps is due entirely, to the much higher proportion of childless couples in this group.

On the basis of the information collected in the Survey all of the couples in the sample met the demographic, religious, and educational requirements for inclusion in the detailed Study. Nevertheless, among the 1,545 couples from whom the information called for in the first interview of the Study was obtained,²⁹ 234, or 15.1 per cent, were found to be ineligible, in most cases because of incorrect entries on the Household Survey schedules.³⁰ If the Survey had functioned perfectly in locating couples for the Study the sample would not have contained these 234 couples but instead would have contained other couples (probably numbering between ninety-six and 114) who actually were eligible but who appeared to be ineligible because of errors in the Survey information. In most cases the misclassification of an eligible couple as ineligible, or *vice versa*, occurred because either (a) the Survey canvasser could not find a member of the household at home and obtained erroneous information from a neighbor, or (b) the wife, husband, or relative was unable or unwilling to answer the questions correctly.³¹ Having an ineligible couple listed as eligible wasted the interviewers' time, but could not bias the group of eligible couples for whom schedules were completed. In contrast, the listing of an eligible couple as

²⁹ The 1,545 couples are those remaining after excluding from the 1,865 in Sample B the 217 "not called on" and the 103 "called on but unknown as to eligibility."

³⁰ Incorrect entries were found on 224 schedules, and were discussed by the authors in Social and Psychological Factors Affecting Fertility, III. The Completeness and Accuracy of the Household Survey of Indianapolis. The Milbank Memorial Fund Quarterly, July, 1945, xxiii, No. 3, pp. 254-296. (Reprint, pp. 95-137.) In addition, seven couples were ineligible because the marriage was broken by separation, divorce, or death between the Survey and the Study, and three because age of wife at marriage appeared to be under 30 when computed from the data on the Survey schedules (current age in years minus the difference between 1941 and year of marriage) but actually was 30 or older.

³¹ In about half of these cases year of marriage was reported incorrectly.

ineligible for either of the reasons just mentioned could introduce a slight bias, for example, the under-representation of couples difficult to find at home (discussed above), and of those lacking in cooperation (discussed below). It is believed, however, that errors in the Survey data were too infrequent or too small to affect significantly the representativeness of the couples interviewed.

Of the 1,648 couples in Sample B who were seen by the interviewers (i.e., an interviewer talked with some member of the household), 147 refused to cooperate in the Study. Four would not allow the interviewers to explain the nature and purpose of the Study, sixty-six listened to an explanation but would do nothing more, and seventy-seven answered the questions on Form A (the relatively short form used in the first interview) but would not answer those on the other forms. In addition, 417 couples were not classified as to cooperation.²⁵ Deducting these from the 1,648 couples who were seen by the interviewers and relating the remainder to the 147 who would not cooperate gives a refusal rate of 11.9 per cent. This percentage is too large, however, for it is almost certain that several of the seventy couples who would not answer any question would not have met the eligibility requirements. Allowing for them on a proportional basis reduces the refusal rate to 11.1 per cent.²⁶

²⁵ The main reasons are as follows: (a) 234 couples were found to be ineligible, hence there was no reason to record an opinion as to their probable cooperation; (b) in accordance with sampling plan B-1 discussed earlier, 107 childless couples classified as relatively sterile were asked only the questions on Form A and might or might not have answered those on Form S; and (c) most of the remaining seventy-six couples were not seen until shortly before the date when the interviewer stopped working in their part of the City or the date when the field work ended, and the interviewer did not determine whether they would cooperate.

²⁶ In the opinion of the writers, the fact that nearly 90 per cent of the couples agreed to cooperate in the Study is due primarily to three causes. One is the high quality of the staff of interviewers, which included Mrs. Martha Sampson Herrick and Mrs. Emily Marks Skolnick, Supervisors, and Mary M. Aikin, Miriam Bintz, Frances N. Butts, Margaret Creviston, Getrude D. Davis, Vida Davison, Dorothy McMillin Gross, Helen Jennings, Margaret A. McConnell, Ruth G. Moss, and Virginia Kahn White. Mrs. Gross and Miss Jennings had the lowest refusal rates, the former completing schedules for ninety-eight couples with four refusals, and the latter for 108 couples with five refusals.

Another factor encouraging cooperation was the aid received from the Indianapolis Committee on American Family Life. This Committee consisted of prominent citizens who were willing to sponsor the Study, namely, Rev. Harry E. Campbell, Alex E. Gordon, Mrs.

(Continued on page 65)

With a few couples the refusal to cooperate was blunt and rude; with the majority it was definite but gracious; with some it was expressed by not finding a convenient time for the second interview, by making appointments but failing to keep them, and by other means of "stalling." In nearly half of the refusal cases (sixty-seven of 147) the wife was not cooperative, hence the interviewers did not try to interest the husband in the Study. Occasionally the husband was present and stimulated or seconded the wife's refusal. If the wife said she would answer the questions but was sure her husband would not, attempts were made to see the husband. The interviewers found that some wives were mistaken as to their husbands' attitudes, and that others were claiming a noncooperative husband merely because of their own unwillingness to participate in the Study, for several of the husbands who were seen in such cases proved willing to answer the questions. Unfortunately the interviewers were unable for various reasons to see thirty-one husbands reported as uncooperative; hence these couples were classified on the basis of the wife's statement.

If respondents would not cooperate the interviewer attempted to ascertain their reasons. Eleven wives and seven husbands said that the questions were too personal (the expressions varying from a polite statement to a brisk "none of your damn business"). Five of these wives judged the questions partly or wholly on the basis of talks with relatives or friends who had been seen previously by an interviewer. (Six other wives who refused were influenced by similar conversations but did not state the reasons involved.) Oppo-

Benjamin D. Hitz, Ralph W. Husted, Hugh McK. Landon, Thurman B. Rice, M.D., and Daniel S. Robinson.

Finally, it is believed that paying the couples who cooperated (one dollar each to the wife and husband if the couple was classified as relatively fecund and supplied the data for five forms, and fifty cents to the wife if the couple was classified as relatively sterile and answered the questions on two forms), contributed appreciably to the low refusal rate. Although this provision appeared to be the decisive argument with relatively few couples, it undoubtedly influenced many. Perhaps its greatest contribution, however, was the stimulation of morale among the interviewers through making them feel that the Study was not a complete imposition on the respondents since they were receiving something for their time and trouble.

sition to giving opinions or answering questions (regardless of how personal) was expressed by eight wives and six husbands. "Too busy" was the excuse of thirteen wives and six husbands, but it appeared to be justified by long working hours or an unusually large amount of family duties for only four wives and three husbands. "I don't want to be bothered," "I'm not interested," "It's a useless study," and similar comments were recorded for sixteen wives and eleven husbands. Remarks of a few other types were made by two or three couples, but for most of the remainder none was recorded.

The important question here, of course, is whether the couples who refused to cooperate differ from the other couples with respect to the characteristics under investigation in the Study. A clean-cut comparison of the two groups cannot be made, however, because seventy couples who refused to answer any question cannot be classified as to eligibility. For this reason it is desirable to compare the uncooperative group not only with the group composed primarily of cooperative couples (all of whom were eligible) but also with this group plus the ineligible couples. As before, the comparisons must be limited to the items on the Survey schedules. The largest absolute difference occurs in the percentage of husbands born in Indiana, which is 66.9 for the uncooperative couples and 71.4 for the cooperative couples. (*See Table 7.*) The largest relative difference occurs in the average number of live births, which is 1.6 for the uncooperative couples and 1.8 for the other two groups. Neither difference is statistically significant. It is possible, but not probable, that more important differences would be found if the number of couples were sufficiently large to permit comparisons within each parity. It is quite probable, however, that some larger differences would be found between cooperative and noncooperative couples if similar comparisons could be made for all items on the Study schedules.

There remain to be considered eighty couples whose schedules

were not completed for reasons not yet discussed. A few of them could not be interviewed because of defective sight or hearing. A few others were excluded because all of their children had died—a situation for which no provision was made in the Study schedules. Several of the eighty couples had good reasons for not cooperating in the Study when first seen by an interviewer (e.g., sickness, husband out of City or working unusually long hours) but were believed willing to do so later. They were not seen after the reason ceased to apply, either because the interviewer's work took her to other parts of the City, or because the field work ended before the delaying conditions changed. The remaining couples were seen

Table 7. Characteristics of couples in Sample B¹ who were seen by an interviewer, and who were (a) uncooperative, (b) cooperative, and (c) cooperative or ineligible.

CHARACTERISTICS	UNCO-OPERATIVE COUPLES ²	COOPERATIVE COUPLES ³	COOPERATIVE OR INELIGIBLE COUPLES ⁴
Number of Couples	147	1,234	1,468
Average:			
Number of Live Births	1.6	1.8	1.8
Age (Wife)	34.7	33.9	34.0
Age (Husband)	37.8	36.7	36.9
Date of Marriage ⁵	7-2-28	7-31-28	7-27-28
Highest Grade of School			
Completed ⁶ (Wife)	11.2	11.2	11.1
Completed ⁶ (Husband)	11.2	11.2	11.1
Rental Value of Dwelling Unit	\$37.21	\$35.37	\$35.02
Per Cent Owning Homes	40.1	43.6	42.3
Per Cent Born in Indiana (Wife)	71.4	72.6	71.8
Per Cent Born in Indiana (Husband)	66.9	71.4	70.7

¹ See Table 5, footnote 1.

² Includes some couples (probably between 8 and 12) who would have been classified as ineligible if answers could have been obtained to the questions on Form A. The number of "unknowns" in the Survey is four for monthly rent or rental value, and zero, one or two for the other items.

³ Includes 150 couples classified "unknown" as to cooperation, of whom 107 are "sterile deferred", i.e., three of every four of the couples with no live birth called on after September 20 and classified as sterile, who were asked only the questions on Form A. It is probable that three to five of them would not have cooperated, i.e., would not have answered the questions on Form S. Most of the other 43 couples "unknown" as to cooperation were not called on until shortly before the end of the field work in their district or in the entire City. It is probable that three to five of them would not have cooperated even though the field work had been continued a few weeks longer. The number of "unknowns" in the Survey is 30 for monthly rent or rental value, and zero to three for the other items.

⁴ Includes the 150 couples described in 3. The number of "unknowns" in the Survey is 40 for monthly rent or rental value, and zero to four for the other items.

⁵ See Table 5, footnote 4.

⁶ See Table 5, footnote 5.

for the first time when the field work was nearly over. If the couples composing the last two groups had been revisited later, it is probable that a few would have been found ineligible, a few would have refused to cooperate, but that schedules would have been completed for a large majority.

As would be expected, the expansion of the sampling ratio for childless couples from 75 to 100 per cent less than a month before the field work ended resulted in a large proportion of childless couples in the group seen by an interviewer but whose schedules were not completed for the reasons listed in the preceding paragraph. In consequence, the average number of live births to the group (1.2) is well below that for all other couples (1.7). In contrast, the differences between the two groups with respect to other

Table 8. Characteristics of (a) couples who were seen but whose schedules were not completed for miscellaneous reasons, and (b) other couples in Sample B.¹

CHARACTERISTIC	COPLES WHO WERE SEEN BUT WHOSE SCHEDULES WERE NOT COMPLETED FOR MISCELLANEOUS REASONS ²	OTHER COUPLES IN SAMPLE B
Number of Couples	808	1,785 ⁴
Average:		
Number of Live Births	1.2	1.7
Age { Wife	34.4	34.1
Husband	37.6	37.0
Date of Marriage ⁵	8-17-48	7-26-48
Highest Grade of School		
Completed ⁶ { Wife	11.2	11.1
Husband	11.3	11.1
Rental Value of Dwelling Unit	\$37.30	\$35.56
Per Cent Owning Homes	42.5	40.6
Per Cent Born in Indiana { Wife	75.0	71.2
Husband	70.9	69.7

¹ See Table 5, footnote 1.

² These reasons are listed in the text.

³ The number of "unknowns" in the Survey is two for age of husband, and zero or one for the other items.

⁴ The number of "unknowns" in the Survey is 53 for monthly rent or rental value, and one to six for the other items.

⁵ See Table 5, footnote 4.

⁶ See Table 5, footnote 5.

characteristics recorded on the Survey schedules are small and not important statistically. (*See Table 8.*) It is probable, therefore, that the Study is not biased appreciably because of the selection just described.

C. ADJUSTING FOR SAMPLING

When the interviewing ended on January 31, 1942, the "study status" of the 2,589 couples, by number of live births reported in the Household Survey, was that shown in Table 9. The first point to be considered with respect to this table is the status of the "deferred" sterile couples. Tables 1 through 8 refer to Sample A or B, both of which include the "deferred" sterile couples because they were assigned to interviewers and answered the questions on Form A. In contrast, Table 9 relates to Sample B-1, which does not include the "deferred" sterile couples who were not asked the more numerous and detailed questions on Form S. This change in reference is necessary because of the procedures connected with the adjustment for sampling.

As shown in Tables 9 and 10, schedules had been completed for 860 fecund couples and for 220 sterile couples. For reasons described in the preceding sections, the percentage distribution by parity of the total 1,080 couples differs markedly from that of the 2,589 couples composing the original universe of eligible couples. (*See Table 11, lines 1 and 3.*) These differences are due mainly to the sampling procedures by parity, and introduce the need for proper weighting of any rates or averages for two or more parities combined that are computed in the analysis of the data. In other words, there is the necessity of adjusting for sampling.

In order to determine the weights that should be used for each parity in this adjustment for sampling, it was necessary to consider whether the couples for whom schedules were completed could justifiably be assumed to be typical of the other couples of the same parity, and if so, for how many of them. Each of the categories in

Table 9. Study status, at termination of field work, of the 2,589 couples in the original universe and of the 1,758 couples in Sample B-1 by number of live births reported in the Household Survey.

STUDY STATUS AT TERMINATION OF FIELD WORK		TOTAL	NO. LIVE BIRTHS	ONE LIVE BIRTH	TWO LIVE BIRTHS	THREE LIVE BIRTHS	FOUR OR MORE LIVE BIRTHS
ORIGINAL UNIVERSE		2,589*	529	727	801	310	221
Not In Sample B-1		831	104	330	397	0	0
Not Seen		724	0	329	395	0	0
Seen, Relatively Sterile, "Deferred"		107	104	1*	2*	0	0
In Sample B-1		1,758	425	397	404	310	221
Not Seen		217	116	34	35	19	12
Seen		1,541	309	363	369	205	209
Unknown Eligibility and Fecundity		103	30	24	19	16	5
Not Cooperative		70	20	20	12	13	5
Unknown Cooperation		33	19	4	7	3	0
Ineligible, Unknown Cooperation		234	64	45	46	37	43
Relatively Fecund		125	9	26	26	22	43
Relatively Sterile		45	28	6	7	4	0
Unknown Fecundity		64	27	13	13	11	0
Eligible		1,204	206	294	304	236	162
Cooperative		1,084	174	260	278	221	151
Relatively Fecund		863	94	183	236	199	151
Schedules Completed		860	93	182	236	199	150
Schedules not Completed		3	1	1	0	0	1
Relatively Sterile		221	80	77	43	28	0
Schedules Completed!		220	80	76	43	28	0
Schedules not Completed		1	0	1	0	0	0
Not Cooperative		77	18	23	17	10	9
Relatively Fecund		67	15	19	15	9	9
Relatively Sterile		8	3	4	1	0	0
Unknown Fecundity		4	0	0	1	1	0
Unknown Cooperation		43	14	11	9	7	2
Relatively Fecund		35	9	8	9	7	2
Relatively Sterile		5	2	3	0	0	0
Unknown Fecundity		3	3	0	0	0	0

* Includes one couple not reporting number of live births in the Household Survey. ** Includes three couples in Sample B-1 who did not report in the Household Survey. *** Includes three couples in Sample B-1 who did not report in the Household Survey. **** Includes three couples in Sample B-1 who did not report in the Household Survey. **** Includes three couples in Sample B-1 who did not report in the Household Survey.

Table 9 was examined in turn. The disposition of the 234 couples who were found to be ineligible for the Study was easily decided. These should be disregarded in determining the weights desired.

Table 10. Extent of adjustment for sampling, by fecundity and number of live births.

LIVE BIRTHS REPORTED IN HOUSEHOLD SURVEY (1)	ELIGIBLE COOPERATIVE COUPLES			WEIGHTS FOR INFLATING (2) + (3) (5)
	Estimated Total Number ¹ (2)	Schedules Completed (3)	Additional Schedules Needed (4)	
RELATIVELY FECUND COUPLES				
TOTAL	1,444	860	584	
0	137	93	44	1.47
1	385	181	203	2.11
2	539	236	303	2.28
3	221	199	22	1.11
4	105	98	7	1.07
5+	57	52	5	1.10
RELATIVELY STERILE COUPLES				
TOTAL	533	110	313	
0	248	80	168	3.10
1	164	76	88	2.16
2	97	41	55	2.31
3	24	22	2	1.09
ALL COUPLES (BY ADDITION)				
TOTAL	1,977	1,080	897	
0	385	173	212	
1	549	258	291	
2	636	278	358	
3	245	221	24	
4	105	98	7	
5+	57	52	5	

¹ Representing numbers desired in the inflated samples of fecund and sterile couples.
See text for description of method of estimating.

A decision regarding the couples who refused to cooperate was not so simple. Some of these couples undoubtedly were ineligible and should be disregarded for that reason. Those who were eligible were known to be uncooperative but there was no way of determining with assurance whether they differed in other respects from the couples for whom schedules were completed. After studying the problem carefully, the Committee conducting the Study concluded

Table 11. Distribution by number of live births reported in the Household Survey, for the original universe of eligible couples and for the inflated and uninflated groups of fecund and sterile couples for whom schedules were completed.

STUDY STATUS	TOTAL	No LIVE BIRTH	One LIVE BIRTH	Two LIVE BIRTHS	Three LIVE BIRTHS	Four or More LIVE BIRTHS
		PERCENTAGE DISTRIBUTION				
Original Universe	100.0	20.4	28.1	31.0	12.0	8.5
<i>Schedules Completed:</i>						
Total—Inflated Sample	100.1	19.5	27.8	31.2	12.4	8.1
Total—Not Inflated	100.0	16.0	23.9	25.7	20.5	13.9
Fecund—Inflated Sample	100.0	9.5	26.7	37.3	15.3	11.2
Fecund—Not Inflated	99.9	10.8	21.2	27.4	23.1	17.4
Sterile—Inflated Sample	100.0	46.5	30.8	18.1	4.5	0.0
Sterile—Not Inflated	100.0	36.4	34.5	19.1	10.0	0.0
NUMBERS						
Original Universe ¹	2,589	529	727	801	310	221
<i>Schedules Completed:</i>						
Total—Inflated Sample	1,977	385	549	636	245	162
Total—Not Inflated	1,080	173	258	278	221	150
Fecund—Inflated Sample	1,444	137	385	539	221	162
Fecund—Not Inflated	860	93	182	236	199	150
Sterile—Inflated Sample	533	248	164	97	24	0
Sterile—Not Inflated	220	80	76	41	22	0

¹ Includes one couple not reporting number of live births in the Household Survey.

that couples refusing to cooperate in a Study of Family Life (the title used by the interviewers in talking to couples) might reasonably be considered as more self-centered than cooperative couples, and as differing from them in matters related to the heart of the Study, particularly in attitudes toward size of family. It seemed advisable, therefore, to disregard such couples in determining the weights to be applied to those for whom schedules were completed.

The next categories examined in Table 9 included the couples unknown as to eligibility or cooperation or both. The majority (831) of these couples were not in Sample B-1 and consisted of two subgroups, namely, 724 couples who were not assigned to interviewers and 107 who were sterile "deferred" and hence unknown as to cooperation. The next largest group consisted of 217 couples in Sample B-1 who were assigned to interviewers but not called on for reasons discussed previously.²⁴ Finally, there were seventy-six couples who were seen by an interviewer, but who nevertheless could not be classified as to eligibility or cooperation. In the opinion of the Committee, it seemed reasonable to assume that (a) these groups contained eligible couples for whom schedules could have been completed if the field work had been continued longer; (b) the couples referred to in "a" were as numerous relatively within the "unknown" groups as the "known" groups; and (c) the couples for whom schedules were completed were typical of those in "a."

In addition to the foregoing categories in Table 9 it was necessary to consider a category not in the table, namely, the couples who actually were eligible but who appeared to be ineligible because of incorrect entries on the Survey schedules. As indicated in an earlier article, such couples are believed to number between 96 and 114.²⁵ If they had been visited by an interviewer it is probable that a large

²⁴ Had the interviewers seen and filled out Form A for these couples, some would have been classified as "deferred" sterile childless couples and hence excluded from Sample B-1.

²⁵ Whelpton, P. K. and Kiser, Clyde V.: Social and Psychological Factors Affecting Fertility. III. The Completeness and Accuracy of the Household Survey of Indianapolis. The Milbank Memorial Fund *Quarterly*, July, 1945, xxiii, No. 3, p. 296. (Reprint, p. 137.)

majority would have cooperated, and that some would have refused. In theory it would have been correct to assume that the cooperative couples in the group are like the couples for whom schedules were completed. From a practical standpoint, however, it seemed wise to disregard the "apparently ineligible but actually eligible" couples when determining weights because nothing was known about their characteristics and the estimated number of these couples was so small that the weights would be approximately the same if they were included or excluded.

The final problem considered in connection with weighting related to the classification by fecundity. Because some of the couples not in Sample B-1 were known to be relatively sterile (the childless "deferred" couples) it was necessary to obtain separate weights for the relatively fecund and relatively sterile couples. To do so it was assumed that the proportions of relatively fecund and sterile couples among those classified "unknown as to fecundity" were the same as they were among those classified as of known fecundity.

The foregoing considerations appeared to warrant the assumption that the relatively fecund couples for whom schedules were completed are typical of the estimated total number of eligible cooperative fecund couples. Hence it was decided to use the latter group as the basis for adjusting the former to allow for sampling by parity. A corresponding assumption and decision were made with respect to the relatively sterile couples. To estimate the total numbers of eligible cooperative couples (fecund and sterile separately), the procedure outlined below was followed within each parity.

1. Couples classed as "not in Sample B-1 and not seen," "in Sample B-1 but not seen," and "seen but unknown as to eligibility, fecundity, and cooperation" were distributed proportionally among the remaining study-status classes.

2. The percentages of couples classified as relatively fecund and relatively sterile were computed on the basis of those of known fecundity.

3. Within each "eligible" study-status class the couples of "unknown

fecundity" were assigned to fecund and sterile groups on the basis of percentages derived in step 2.

4. Within the eligible fecund and sterile groups separately, couples classed as "unknown as to cooperation"⁷⁸ were assigned proportionately to the "cooperative" and "not cooperative" classes.

5. In steps 1 and 4, whenever cases of "unknowns" were logically assigned either to the fecund or sterile "schedules completed" group the actual assignments were made to the "schedules not completed" group. There were thus derived for each parity the adjusted numbers of "eligible, cooperative, fecund" and "eligible, cooperative, sterile" couples for whom schedules were not completed, but presumably would have been completed if all eligible couples in the universe had been interviewed. Adding these numbers to the corresponding numbers of couples for whom schedules were completed and dividing the totals by the latter gave the weights desired. (See Table 10.)

Weights of the type described above could be used in either of two ways: (a) the component rates or averages for the several parities could be computed separately and weighted mathematically each time a composite rate is desired; or (b) the punch cards could be "inflated" by duplicating for each parity the number of cards indicated by the weights, and the rates or averages computed directly. The latter plan was adopted. The numbers of fecund and sterile couples for whom schedules were completed and the numbers of additional cases needed are shown by parity in Table 10. Thus, for fecund childless couples, forty-four cases were needed in addition to the ninety-three for whom schedules were completed to yield a total of 137 in the inflated group. The problem, therefore, was that of choosing at random forty-four cases from the group of ninety-three and making one duplicate punch card for each of them. For fecund couples with one live birth, it was necessary to duplicate 161 cases once and twenty-one twice in order to bring the inflated group to 385.

⁷⁸ Nearly all of the sterile couples unknown as to cooperation were those "deferred" and classified as "not in Sample B-1" in Table 9.

⁷⁹ With this procedure one might use as weights the percentage distribution by parity of the estimated total number of eligible cooperative fecund (or sterile) couples.

To minimize the possibility of bias in the selection of punch cards to be duplicated, Tippett's random numbers were utilized. The 860 punch cards for "fecund couples, schedules completed" were classified by the number of live births reported in the Household Survey.⁷ Tippett's random numbers⁸ were punched on the cards, beginning with the pack of cards for the childless couples. After all parities were completed, the punch cards for each parity were arranged in order by ascending random number. The process was then simply that of taking for duplication the desired number of cards with lowest random numbers. Thus, of the ninety-three cards for childless fecund couples, the forty-four cards with lowest random numbers were selected for duplication. A similar process was used for the remaining parities of fecund and sterile couples.⁹

D. THE REPRESENTATIVENESS OF THE INFLATED SAMPLE

The representativeness of the inflated groups may be considered from several points of view. Since the purpose of inflation is to secure representativeness by number of children ever born, the extent to which this objective is attained may first be considered. As indicated in Table 11, lines 1 and 2, when the inflated samples for the fecund and sterile couples are combined, the percentage distribution by parity is very much the same as that observed for the original universe of eligible couples. In spite of the foregoing, one may ask what evidence there is that the two inflated groups are

⁷ In a few instances the number of live births reported in the Household Survey differed from the number reported in the intensive interviews of the Study. The former were used for purposes of the inflation, however, since they had formed the basis for the original sampling ratios, and were the only data on live births for couples not seen by interviewers.

⁸ There are twenty-six pages in Tippett's booklet, each page containing eight columns and each column fifty four-digit numbers. The first column numbers were used in order. Hence, for the 860 cards for fecund couples, the numbers drawn from Tippett are those in the first columns of pages 1-17 and part of those in the first column of page 18. For the 220 sterile couples the entire first columns of pages 19-21 and parts of those of pages 18 and 22 were utilized. See Tippett, L. H. C.: *RANDOM SAMPLING NUMBERS*. London, Cambridge University Press, 1927, 26 pp.

⁹ The mechanical work of duplicating the cards was done in the statistical office of the School of Hygiene and Public Health, The Johns Hopkins University, through the courtesy of Professor Lowell J. Reed, Chairman of The Committee on the Study of Social and Psychological Factors Affecting Fertility.

representative of the respective fecund and sterile parts of the universe with respect to distributions by parity. It will be recalled that the numerical distributions by parity of the inflated groups of fecund and sterile couples are equivalent to the numbers estimated for fecund and sterile "eligible cooperative" couples. These were obtained by utilizing the detailed sampling status of the total universe for purposes of allocating the couples "not in Sample B-1," "in Sample B-1 but not seen," and "seen but unknown as to fecundity, eligibility, and cooperativeness." Hence, to the extent that no biases by parity are introduced by the exclusion of ineligible, apparently ineligible but actually eligible, and noncooperative couples, the distribution by parity of the inflated groups of relatively fecund and sterile couples combined should resemble that for the original universe of 2,589 couples. The fact that this similarity does exist supports the belief that the distributions by parity in the inflated groups of fecund and sterile couples separately are substantially correct.

The above considerations refer to the representativeness achieved by the proper *amount* of inflation for each parity of fecund and sterile couples. A question still remains concerning representativeness within each parity. This is a function of the *method* of selecting the punch cards that were to be duplicated. As previously stated, Tippett's random numbers were used in the hope that the inflated group for a given parity would not depart significantly from the uninflated group in so far as characteristics of the couples are concerned. That this objective was achieved may be seen in the similarity of the inflated and uninflated groups of fecund couples and in the similarity of the inflated and uninflated sterile groups with respect to the descriptive characteristics considered in Tables 12-18. These characteristics relate to age, state of birth, and education of the husband and wife; year of marriage; and tenure and rental value of the dwelling unit.

By way of illustration, a few comparisons may be made between

the inflated and uninflated groups of relatively fecund one-child couples. Thus, the median age of husband is 36.5 years in the inflated, and 36.6 years in the uninflated group. The median age of wife is precisely the same, 34.3, in the two groups. (See Tables 12-13.) The proportions of husbands and wives born in Indiana, in other northern areas, and in the South are essentially the same for the inflated as for the uninflated groups. (See Table 14.) The median grade in school completed by husband is precisely the same, 12.2, in both groups. The median grade completed by wife is 12.2 in the inflated group and 12.3 in the uninflated group. (See Tables 15-16.) The distributions of couples by year of marriage and tenure of the home are virtually the same in the inflated and uninflated groups. The median rent of dwelling unit (\$34.44 and \$34.49) is also almost exactly the same for the two groups.

The comparison of the inflated with the uninflated groups, of course, is no rigorous test of Tippett's random numbers. Suffice it to say, however, that other students have carried out special tests of Tippett's numbers and have reported favorably on their random character. In the present instance the important point is that Tippett's numbers appeared to "fill the bill" quite well. The inflated groups are quite similar to the uninflated groups of given parities among both fecund and sterile couples.

Furthermore, within each parity the inflated groups of fecund and sterile couples combined are fairly representative of the original universe of eligible couples. This may be seen by comparing the two top lines under each parity in Tables 12-18. Several figures may again be cited for one-child couples by way of illustration. Thus, the median age of husband is 36.5 in the "total inflated" groups and 36.9 in the "original universe." For median age of wife the respective figures are 34.5 and 34.6. The median highest grade in school completed is 12.2 for the husband and also for the wife in each of the two groups. The median monthly rental is \$34.44 for the "total inflated" group and \$35.05 for the "original universe." Equally close

Table 12. Median age and percentage distribution by age of husband for couples of given study status and number of live births.

STUDY STATUS	ME-DIAN AGE	PERCENTAGE DISTRIBUTION BY AGE OF HUSBAND								
		29-31	32-33	34	35	36-37	38-39	40-41	42-44	45-54
NO LIVE BIRTH										
Original Universe	36.1	1.9	10.6	9.1	13.1	18.8	15.4	14.6	8.3	8.2
Total Inflated Sample	37.7	1.3	13.0	10.4	14.5	18.7	13.2	17.4	5.7	5.7
Fecund—Schedules Completed	37.8	1.5	8.8	11.7	12.4	23.4	15.3	14.6	7.3	5.1
Inflated Sample	37.7	2.2	9.7	11.8	10.8	25.8	14.0	11.8	7.5	6.5
Not Inflated										
Berile—Schedules Completed	37.5	1.2	15.3	9.7	15.7	16.1	12.1	19.0	4.8	6.0
Inflated Sample	37.6	1.3	15.0	10.0	15.0	16.3	12.5	18.8	5.0	6.3
Not Inflated										
ONE LIVE BIRTH										
Original Universe	36.9	7.3	13.7	11.3	13.7	22.3	14.9	6.2	6.4	5.1
Total Inflated Sample	36.5	9.8	14.2	14.6	11.3	19.7	15.8	2.7	5.5	6.4
Fecund—Schedules Completed	36.5	9.6	16.4	11.2	12.5	19.5	15.6	3.1	4.9	7.3
Inflated Sample	36.6	9.3	15.9	11.5	12.6	19.2	15.9	3.3	4.9	7.1
Not Inflated										
Berile—Schedules Completed	36.4	10.4	9.1	22.6	8.5	20.1	16.5	1.8	6.7	4.3
Inflated Sample	36.5	10.5	9.2	21.1	9.2	21.1	17.1	1.3	6.6	3.9
Not Inflated										
TWO LIVE BIRTHS										
Original Universe	36.4	9.7	16.6	11.9	13.6	18.7	12.9	8.9	5.2	2.5
Total Inflated Sample	36.2	10.1	18.1	11.9	14.3	19.3	12.1	8.8	3.5	1.9
Fecund—Schedules Completed	36.2	8.5	18.2	12.8	15.2	19.3	11.9	8.5	3.3	2.2
Inflated Sample	36.2	9.3	18.2	12.3	15.3	19.5	11.4	8.9	3.0	2.1
Not Inflated										
Berile—Schedules Completed	36.2	18.6	17.5	7.2	9.3	19.6	13.4	10.3	4.1	0.0
Inflated Sample	36.0	19.0	19.0	7.1	9.5	19.0	11.9	9.5	4.8	0.0
Not Inflated										
THREE LIVE BIRTHS										
Original Universe	36.6	8.7	15.9	12.3	12.3	21.7	13.9	9.4	3.9	1.9
Total Inflated Sample	36.5	9.0	16.8	11.5	12.3	23.4	12.3	9.8	3.7	1.2
Fecund—Schedules Completed ¹	36.5	8.6	17.7	11.8	12.3	22.3	12.7	9.5	3.6	1.4
Inflated Sample	36.5	8.6	16.7	12.1	12.1	22.7	13.6	9.1	3.5	1.5
Not Inflated										
FOUR OR MORE LIVE BIRTHS										
Original Universe	34.8	17.6	29.0	11.3	8.6	15.4	9.5	4.5	3.6	0.5
Total (or Fecund) Inflated Sample	34.8	14.8	30.9	14.2	9.3	13.0	10.5	3.7	3.7	0.0
Total (or Fecund) Not Inflated	34.9	15.3	30.0	13.3	10.0	13.3	10.7	4.0	3.3	0.0

¹Distributions are not shown for the relatively sterile couples with three live births, since there were only 22 such couples (24 in the inflated group).

Table 13. Median age and percentage distribution by age of wife for couples of given study status and number of live births.

STUDY STATUS	ME-DIAN AGE 30	PERCENTAGE DISTRIBUTION BY AGE OF WIFE							
		Under 30	30-31	32-33	34	35	36-37	38-39	40-44
NO LIVE BIRTH									
Original Universe	35.7	4.7	14.0	17.8	10.6	12.5	16.1	10.0	14.4
Total Inflated Sample	35.5	8.1	13.8	16.6	11.9	12.2	19.5	6.8	11.1
Fecund—Schedules Completed									
Inflated Sample	35.2	0.7	16.1	25.5	11.7	10.2	19.7	5.8	10.1
Not Inflated	35.2	1.1	16.1	24.7	11.8	9.7	20.4	5.4	10.8
Sterile—Schedules Completed									
Inflated Sample	35.6	12.1	12.5	11.7	12.1	13.3	19.4	7.3	11.7
Not Inflated	35.7	11.3	12.5	11.3	12.5	13.8	20.0	7.5	11.3
ONE LIVE BIRTH									
Original Universe	34.6	6.9	19.2	22.3	11.9	13.1	12.6	8.3	5.8
Total Inflated Sample	34.5	7.3	19.3	24.0	13.3	11.5	12.4	9.3	2.9
Fecund—Schedules Completed									
Inflated Sample	34.3	7.5	20.8	24.7	13.5	9.9	11.4	9.6	2.6
Not Inflated	34.3	7.1	20.3	25.3	13.7	9.9	11.5	9.3	2.7
Sterile—Schedules Completed									
Inflated Sample	34.9	6.7	15.9	22.6	12.8	15.2	14.6	8.5	3.7
Not Inflated	34.8	6.6	15.8	23.7	13.2	14.5	14.5	7.9	3.9
TWO LIVE BIRTHS									
Original Universe	33.9	11.6	20.5	25.8	9.9	9.4	11.6	7.4	3.9
Total Inflated Sample	33.4	13.5	25.0	25.6	8.8	6.8	12.1	6.3	1.9
Fecund—Schedules Completed									
Inflated Sample	33.5	11.3	26.9	24.7	8.9	6.9	12.6	6.5	2.1
Not Inflated	33.4	11.9	27.1	24.6	8.9	6.8	12.3	6.4	2.1
Sterile—Schedules Completed									
Inflated Sample	33.1	25.8	14.4	30.9	8.2	6.2	9.3	5.2	0.0
Not Inflated	33.1	26.2	14.3	31.0	7.1	7.1	9.5	4.8	0.0
THREE LIVE BIRTHS									
Original Universe	33.5	11.7	26.3	24.7	10.4	8.8	11.0	3.6	3.6
Total Inflated Sample	33.6	11.5	23.4	28.3	10.2	7.4	11.9	3.3	4.1
Fecund—Schedules Completed ¹									
Inflated Sample	33.6	11.8	22.7	27.3	10.9	7.7	12.3	3.6	3.6
Not Inflated	33.7	11.6	22.3	27.8	12.1	8.1	11.1	3.5	3.5
FOUR OR MORE LIVE BIRTHS									
Original Universe	32.3	22.2	31.7	21.3	9.5	6.3	5.4	2.3	1.4
Total (or Fecund) Inflated Sample	32.3	20.4	33.3	21.0	9.3	5.6	8.0	1.9	0.6
Total (or Fecund) Not Inflated	32.3	20.0	32.7	22.0	8.7	6.0	8.0	2.0	0.7

¹ See footnote 1 to Table 12.

Table 14. Percentage distribution by birth region of the husband and wife for couples of given study status and number of live births.

STUDY STATUS	PER CENT OF HUSBANDS BORN IN			PER CENT OF WIVES BORN IN		
	Indiana	Other Northern States	Southern States	Indiana	Other Northern States	Southern States
NO LIVE BIRTH						
Original Universe	68.8	19.8	11.4	69.1	19.0	11.0
Total Inflated Sample	71.6	20.1	8.3	70.4	20.8	8.8
Fecund—Schedules Completed	70.6	17.6	11.8	71.3	11.9	5.8
Inflated Sample	69.6	18.5	12.0	72.0	11.5	6.5
Not Inflated	72.2	11.4	6.5	69.4	20.2	10.5
Sterile—Schedules Completed	72.5	11.3	6.3	70.0	20.0	10.0
Inflated Sample						
Not Inflated						
ONE LIVE BIRTH						
Original Universe	69.8	17.9	12.3	70.9	18.5	10.6
Total Inflated Sample	73.9	16.6	9.5	69.8	20.0	10.1
Fecund—Schedules Completed	75.8	13.8	10.4	68.8	19.7	11.4
Inflated Sample	75.8	14.3	9.9	68.7	19.8	11.5
Not Inflated	69.1	23.5	7.4	72.0	20.7	7.3
Sterile—Schedules Completed	69.3	24.0	6.7	71.4	21.1	6.6
Inflated Sample						
Not Inflated						
TWO LIVE BIRTHS						
Original Universe	70.8	17.1	11.0	71.4	15.7	11.9
Total Inflated Sample	70.4	17.6	11.9	73.9	14.6	11.5
Fecund—Schedules Completed	71.1	17.8	11.1	73.3	15.6	11.1
Inflated Sample	70.3	18.1	11.4	73.3	15.7	11.0
Not Inflated	67.0	16.5	16.5	77.3	9.3	13.4
Sterile—Schedules Completed	66.7	14.3	19.0	76.2	9.5	14.3
Inflated Sample						
Not Inflated						
THREE LIVE BIRTHS						
Original Universe	70.5	16.6	13.0	72.6	17.3	10.1
Total Inflated Sample	72.0	16.0	11.9	71.9	19.4	8.7
Fecund—Schedules Completed ¹	72.1	16.9	11.0	71.6	19.3	9.2
Inflated Sample	72.1	17.3	10.7	70.9	19.9	9.2
Not Inflated						
FOUR LIVE BIRTHS						
Original Universe	68.3	14.5	17.2	73.8	13.6	12.7
Total (or Fecund) Inflated Sample	72.2	13.0	14.8	75.9	11.7	12.3
Total (or Fecund) Not Inflated	71.3	13.3	15.3	76.0	12.0	12.0

¹ See footnote 1 to Table 12.

Table 15. Median grade and percentage distribution by educational attainment of the husband for couples of given study status and number of live births.

STUDY STATUS	ME-DIAN GRADE	PERCENTAGE DISTRIBUTION BY HIGHEST GRADE OF SCHOOL COMPLETED BY HUSBAND										
		G. S. 8	High School					College				
			1	2	3	4	"Unk."	1-3	4+	"Unk."		
NO LIVE BIRTH												
Original Universe	12.2	22.3	4.2	11.9	5.9	31.6	1.3	9.8	12.5	0.4	Original U	
Total Inflated Sample	12.1	26.5	5.5	9.6	5.2	28.1	2.1	8.8	14.3	0.0	Total Infla	
Fecund—Schedules Completed	12.3	19.0	4.4	11.7	5.1	29.2	3.6	11.7	15.3	0.0	Fecund—S	
Inflated Sample	12.3	17.2	4.3	11.8	5.4	31.2	3.2	10.8	16.2	0.0	Inflated	
Not Inflated											Not Infla	
Sterile—Schedules Completed	11.8	30.6	6.0	8.5	5.2	27.4	1.2	7.3	13.7	0.0	Sterile—Sc	
Inflated Sample	11.9	30.0	6.3	8.8	5.0	27.5	1.3	7.5	13.8	0.0	Inflated	
Not Inflated											Not Infla	
ONE LIVE BIRTH												
Original Universe	12.2	19.9	5.6	10.5	6.1	29.3	1.4	12.5	13.6	1.1	Original U	
Total Inflated Sample	12.2	19.5	5.8	11.7	7.1	27.7	1.3	12.2	14.8	0.0	Total Infla	
Fecund—Schedules Completed	12.2	19.2	4.9	10.9	7.5	30.4	1.8	10.6	14.5	0.0	Fecund—S	
Inflated Sample	12.2	19.8	4.9	10.4	7.1	30.8	1.6	10.4	14.8	0.0	Inflated	
Not Inflated											Not Infla	
Sterile—Schedules Completed	12.1	30.1	7.9	13.4	6.1	21.3	0.0	15.9	15.2	0.0	Sterile—Sc	
Inflated Sample	12.1	21.1	6.6	14.5	5.3	22.4	0.0	14.5	15.8	0.0	Inflated	
Not Inflated											Not Infla	
TWO LIVE BIRTHS												
Original Universe	12.1	21.0	7.3	12.5	5.8	26.4	1.4	11.0	14.6	0.1	Original U	
Total Inflated Sample	12.1	19.8	7.5	14.8	5.0	25.3	0.3	12.1	15.1	0.0	Total Infla	
Fecund—Schedules Completed	12.2	18.4	6.5	15.8	4.3	27.3	0.4	12.6	14.8	0.0	Fecund—S	
Inflated Sample	12.2	18.6	6.4	15.7	4.2	27.5	0.4	12.3	14.8	0.0	Inflated	
Not Inflated											Not Infla	
Sterile—Schedules Completed	10.9	27.8	13.4	9.3	9.3	14.4	0.0	9.3	16.5	0.0	Sterile—Sc	
Inflated Sample	11.0	26.2	14.3	9.5	9.5	14.3	0.0	9.5	16.7	0.0	Inflated	
Not Inflated											Not Infla	
THREE LIVE BIRTHS												
Original Universe	11.2	29.8	4.2	13.6	8.1	26.9	0.3	4.5	12.0	0.0	Original U	
Total Inflated Sample	11.5	25.4	5.7	14.3	7.8	29.1	0.4	5.3	11.1	0.0	Total Infla	
Fecund—Schedules Completed ¹	11.6	25.5	5.9	13.2	8.6	30.0	0.0	4.1	11.8	0.0	Fecund—S	
Inflated Sample	11.6	26.3	5.6	12.1	9.1	29.8	0.0	4.5	11.6	1.0	Inflated	
Not Inflated											Not Infla	
FOUR OR MORE LIVE BIRTHS												
Original Universe	9.8	41.2	10.0	16.7	8.1	14.9	0.9	5.0	3.2	0.0	Original U	
Total (or Fecund) Inflated Sample	10.0	37.7	11.1	15.4	9.9	15.4	1.2	5.6	3.7	0.0	Total (or Fec	
Total (or Fecund) Not Inflated	10.0	38.0	10.7	16.0	10.0	14.7	1.3	6.0	3.3	0.0	Total (or Fec	

¹ See footnote 1 to Table 12.

Table 16. Median grade and percentage distribution by educational attainment of the wife for couples of given study status and number of live births.

STUDY STATUS	ME-DIAN GRADE	PERCENTAGE DISTRIBUTION BY HIGHEST GRADE OF SCHOOL COMPLETED BY WIFE										
		G. S. 8	High School					College				
			1	2	3	4	"Unk."	1-3	4+	"Unk."		
NO LIVE BIRTH												
Original Universe		12.2	18.9	3.8	11.2	4.7	45.3	1.1	6.4	8.1	0.4	
Total Inflated Sample		12.2	17.4	5.2	14.5	5.5	43.4	0.5	4.7	8.8	0.0	
Second—Schedules Completed		12.4	17.5	1.5	9.5	2.2	51.1	1.5	10.9	5.8	0.0	
Inflated Sample		12.4	15.1	1.1	9.7	2.2	52.7	1.1	10.8	7.5	0.0	
Not Inflated												
Sterile—Schedules Completed		12.0	17.3	7.3	17.3	7.3	39.1	0.0	1.2	10.5	0.0	
Inflated Sample		12.0	16.3	7.5	17.5	7.5	40.0	0.0	1.3	10.0	0.0	
Not Inflated												
ONE LIVE BIRTH												
Original Universe		12.2	12.7	7.4	12.2	7.4	39.1	0.8	9.9	10.2	0.3	
Total Inflated Sample		12.2	9.8	8.7	13.7	8.2	40.3	0.0	11.7	7.7	0.0	
Second—Schedules Completed		12.2	6.8	9.6	16.4	7.3	40.8	0.0	12.2	7.0	0.0	
Inflated Sample		12.3	7.1	9.3	15.9	7.1	41.2	0.0	12.1	7.1	0.0	
Not Inflated												
Sterile—Schedules Completed		12.2	17.1	6.7	7.3	10.4	39.0	0.0	10.4	9.1	0.0	
Inflated Sample		12.2	17.1	6.6	6.6	10.5	39.5	0.0	10.5	9.2	0.0	
Not Inflated												
TWO LIVE BIRTHS												
Original Universe		12.2	15.9	7.1	14.0	6.5	36.6	1.1	9.6	9.1	0.1	
Total Inflated Sample		12.2	14.6	7.5	13.8	7.7	37.7	1.1	9.1	8.3	0.0	
Second—Schedules Completed		12.2	14.5	5.8	14.1	8.0	39.0	1.3	9.5	8.0	0.0	
Inflated Sample		12.2	14.4	5.9	14.0	8.1	39.4	1.3	8.9	8.1	0.0	
Not Inflated												
Sterile—Schedules Completed		11.8	15.5	17.5	12.4	6.2	30.9	0.0	7.2	10.3	0.0	
Inflated Sample		11.3	16.7	19.0	11.9	7.1	28.6	0.0	7.1	9.5	0.0	
Not Inflated												
THREE LIVE BIRTHS												
Original Universe		11.4	21.7	6.1	18.1	8.7	29.1	0.6	7.1	8.4	0.0	
Total Inflated Sample		11.6	19.7	5.7	20.1	6.6	31.6	0.8	7.0	8.6	0.0	
Second—Schedules Completed ¹		11.9	19.5	6.4	18.2	6.4	32.3	0.9	7.7	8.6	0.0	
Inflated Sample		11.8	19.7	6.1	18.7	6.6	32.3	0.5	7.6	8.6	0.0	
Not Inflated												
FOUR OR MORE LIVE BIRTHS												
Original Universe		10.4	30.3	12.7	19.9	8.1	24.0	0.0	3.6	1.4	0.0	
Total (or Fecund) Inflated Sample		10.4	27.2	13.6	23.5	8.0	21.6	0.0	4.3	1.9	0.0	
Total (or Fecund) Not Inflated		10.4	28.0	12.7	22.0	8.7	22.7	0.0	4.7	1.3	0.0	

¹ See footnote 1 to Table 12.

Table 17. Percentage distribution by year of marriage and by tenure of home for couples of given study status and number of live births.

STUDY STATUS	YEAR OF MARRIAGE			TENURE OF HOME		
	1927	1928	1929	Owner	Renter	Other
NO LIVE BIRTH						
Original Universe	29.5	30.6	39.9	39.0	57.2	3.8
Total Inflated Sample	28.3	33.5	38.2	43.9	54.3	1.8
Fecund—Schedules Completed	24.1	39.4	36.5	48.9	48.2	2.9
Inflated Sample	24.7	37.6	37.6	49.5	48.4	2.2
Not Inflated						
Sterile—Schedules Completed	30.6	30.2	39.1	41.1	57.7	1.2
Inflated Sample	31.3	31.3	37.5	41.3	57.5	1.3
Not Inflated						
ONE LIVE BIRTH						
Original Universe	29.8	32.0	38.1	45.0	53.3	1.7
Total Inflated Sample	25.9	38.3	35.9	44.1	53.9	2.0
Fecund—Schedules Completed	23.9	35.8	40.3	42.3	54.8	2.9
Inflated Sample	23.1	36.3	40.7	43.4	53.8	2.7
Not Inflated						
Sterile—Schedules Completed	30.5	43.9	25.6	48.2	51.8	0.0
Inflated Sample	30.3	43.4	26.3	47.4	52.6	0.0
Not Inflated						
TWO LIVE BIRTHS						
Original Universe	30.8	30.3	38.8	47.3	51.6	1.1
Total Inflated Sample	30.5	30.3	39.2	52.0	47.3	0.6
Fecund—Schedules Completed	29.7	30.6	39.7	51.8	48.2	0.0
Inflated Sample	29.2	30.5	40.3	51.3	48.7	0.0
Not Inflated						
Sterile—Schedules Completed	35.1	28.9	36.1	53.6	42.3	4.1
Inflated Sample	35.7	28.6	35.7	52.4	42.9	4.8
Not Inflated						
THREE LIVE BIRTHS						
Original Universe	33.9	33.1	32.9	39.1	59.9	1.0
Total Inflated Sample	34.3	32.2	33.5	41.0	58.2	0.8
Fecund—Schedules Completed ¹	33.5	33.0	33.5	41.8	57.3	0.9
Inflated Sample	32.7	33.2	34.2	41.9	57.1	1.0
Not Inflated						
FOUR OR MORE LIVE BIRTHS						
Original Universe	30.8	32.1	37.1	28.5	69.7	1.8
Total (or Fecund) Inflated Sample	30.9	34.6	34.6	30.9	67.3	1.9
Total (or Fecund) Not Inflated	30.7	33.3	36.0	30.7	67.3	2.0

¹ See footnote 1 to Table 12.

Table 18. Median rental value and percentage distribution by rental value of dwelling unit for couples of given study status and number of live births.

STUDY STATUS	MEDIAN RENTAL VALUE	PERCENTAGE DISTRIBUTION BY RENTAL VALUE OF DWELLING UNIT								
		Under \$20	\$20-24	\$25-29	\$30-34	\$35-39	\$40-49	\$50-59	\$60-74	\$75+
NO LIVE BIRTH										
Initial Universe	\$37.07	9.4	6.8	13.0	14.6	17.4	17.2	12.4	6.8	3.4
All Inflated Sample	35.56	12.6	8.2	13.7	13.5	17.3	13.7	10.7	6.6	3.6
and—Schedules Completed	37.22	3.8	9.1	16.7	14.4	13.6	17.4	18.2	3.8	3.0
Inflated Sample	37.50	3.3	7.7	15.6	15.6	15.6	15.6	18.9	4.4	3.3
Not Inflated	34.83	17.7	7.8	12.1	12.9	19.4	11.6	6.5	8.2	3.9
Initial Universe	35.17	16.0	8.0	12.0	13.3	20.0	12.0	6.7	8.0	4.0
ONE LIVE BIRTH										
Initial Universe	35.05	11.4	8.5	15.6	14.4	14.5	13.8	10.7	5.1	6.1
All Inflated Sample	34.44	11.0	8.3	16.0	16.7	14.9	12.6	9.7	5.9	5.0
and—Schedules Completed	34.44	8.6	9.1	15.2	19.3	15.2	11.2	10.7	6.4	4.3
Inflated Sample	34.49	8.5	9.6	14.7	19.2	15.3	11.3	10.2	6.8	4.5
Not Inflated	34.44	16.5	6.1	17.7	11.0	14.0	15.9	7.3	4.9	6.7
Initial Universe	34.38	15.8	6.6	18.4	10.5	14.5	15.8	6.6	5.3	6.6
TWO LIVE BIRTHS										
Initial Universe	33.68	15.1	10.9	14.2	13.4	14.8	11.6	7.8	6.1	6.1
All Inflated Sample	33.59	15.7	9.0	14.8	14.6	17.6	12.1	6.0	5.4	4.8
and—Schedules Completed	34.31	13.0	9.3	14.3	15.5	18.1	13.8	6.3	5.0	4.7
Inflated Sample	33.99	12.8	10.0	14.9	15.7	17.0	14.0	6.4	5.1	4.3
Not Inflated	34.44	31.2	7.5	17.2	9.7	15.1	2.2	4.3	7.5	5.4
Initial Universe	33.57	30.0	7.5	17.5	10.0	15.0	2.5	5.0	7.5	5.0
THREE LIVE BIRTHS										
Initial Universe	28.56	27.4	11.6	15.5	9.6	12.9	12.5	2.0	3.6	5.0
All Inflated Sample	29.03	27.1	10.8	15.0	11.3	13.8	11.7	1.7	4.2	4.6
and—Schedules Completed ¹	28.79	26.9	11.6	15.3	10.6	13.4	13.0	1.4	4.2	3.7
Inflated Sample	29.11	25.8	12.4	14.4	9.8	14.4	13.4	1.5	4.1	4.1
Not Inflated										
FOUR OR MORE LIVE BIRTHS										
Initial Universe	19.76	51.2	14.3	12.9	7.8	5.5	4.6	1.8	0.0	1.8
All (or Fecund) Inflated Sample	19.70	51.6	13.8	11.3	8.2	5.7	3.8	2.5	0.0	3.1
All (or Fecund) Not Inflated	19.93	50.3	14.3	11.6	8.8	5.4	4.1	2.7	0.0	2.7

¹ See footnote 1 to Table 12.

similarities may be observed for the couples with other numbers of children.

The reasons for the foregoing types of representativeness are simply that (a) within a given parity the interviewed couples are like those in the universe (partly because a deliberate stratification by rent was followed in selecting couples for interview), and (b) the foregoing inflation of the groups effects no substantial change in the descriptive characteristics within given parities of fecund or sterile couples and at the same time provides a presumably proper fecund-sterile ratio within each parity.

Since the inflated groups of fecund and sterile couples combined are fairly representative of the "original universe" within each of the several parities and since the total inflated group is representative of the "original universe" by parity itself, it follows that the total inflated group should resemble the "original universe" by descriptive characteristics when all couples, regardless of parity, are considered. The extent to which this holds true may be illustrated by the following:

	"Original Universe"	"Total Inflated Group"
Median Age — Husband	36.8	36.4
Median Age — Wife	34.2	34.0
Median School Grade Completed — Husband	12.0	12.0
Median School Grade Completed — Wife	12.1	12.1
Median Rental Value	\$33.29	\$32.81

E. COMPARISON OF COUPLES IN THE STUDY WITH THOSE IN THE HOUSEHOLD SURVEY

It is of interest to consider briefly the characteristics of the inflated group in relation to a universe larger than that of the 2,589 couples. As explained previously, the 2,589 couples are those eligible for the Study on the basis of information from the Household Survey. These couples were reported in the Survey as having the following characteristics: native white, Protestant, married during

1927-1929, husband under 40 and wife under 30 at marriage, neither spouse previously married, couple resided in a large city most of the time since marriage, and both husband and wife had at least graduated from grammar school.

The question naturally arises concerning the bearing of the above-mentioned selections on other descriptive facets of the Study sample. The best answer that can be made is based on a comparison between a few descriptive items for the couples in the inflated Study group and for all native-white, once-married couples with wife under 45 in the Indianapolis Household Survey.¹¹ First of all, the couples in the Study are heavily concentrated by age as compared with those in the Survey. The Survey included all wives 15-44, regardless of age of husband. Since the Study was confined to couples married during 1927-1929, in which the age at marriage was under 30 for the wife and under 40 for the husband, the resulting age limits (as of 1941) are 26-44 for the wives and 29-54 for the husbands.¹² Actually, over one-half of the wives in the Study fall into the 30-34 age group, since the popular bridal ages are 18-22. There is a corresponding concentration of husbands in the 35-39 age group, owing to the tendency of grooms to be a few years older than brides. The age-concentration of husbands, however, is not quite so marked as that of wives, owing to the less severe restrictions imposed on age of grooms than of brides in so far as eligibility for the Study is concerned.

The Study group is characterized by somewhat higher economic status (as measured by monthly rental value of dwelling unit) than is the Survey. For instance, 9.7 per cent of the Study couples as compared with 6.8 per cent of the Survey couples are in the "\$60 and over" rental class. The percentages of couples in the "under \$20" category are 18.2 and 21.8, respectively. The median rental values

¹¹ Items for the Survey group are taken from Whelpton, P. K. and Kiser, C. V.: Social and Psychological Factors Affecting Fertility. The Milbank Memorial Fund *Quarterly*, July, 1943, xxI, No. 3, pp. 221-280. (Reprint, pp. 1-80.)

¹² The younger age limits assume a minimum age at marriage of 15 for wife and 18 for husband.

are \$32.81 and \$30.64 for the Study and Survey couples, respectively. This somewhat higher economic status of the Study couples may be due chiefly to the restriction of the latter to wives and husbands

who had completed at least the elementary grades and to husbands at least 29 years of age. Contrary to what might be expected, the higher economic status of Study couples than of Survey couples cannot be attributed to the restriction to Protestants, for Catholics constitute

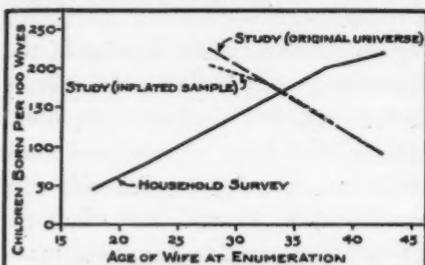


Fig. 1. Number of children ever born per one hundred wives by age of wife, for Protestant couples in (a) the Household Survey; (b) the original universe of couples eligible for the Study; and (c) the inflated Study sample. See Table 19.

the major non-Protestant group in the Indianapolis Household Survey and are characterized by somewhat higher median rental values than Protestants.¹¹

Since the ratio of Protestants to non-Protestants is relatively high in the South, the restriction on religion might be expected to raise the proportion of southern-born persons in the Study. Apparently this restriction is more than offset by the restrictions on education and years lived in a large city since marriage. At all events, only 10.8 per cent of the husbands and 10.3 per cent of the wives in the inflated Study group are southern-born as compared with 13.5 per cent and 12.7 per cent, respectively, in the Household Survey.

Finally, one may ask what bearing the qualifications for eligibility have on fertility. Before making comparisons, however, it is necessary to keep in mind the age and age-at-marriage characteristics of the Study couples. Since the eligible couples were married during 1927-1929, all Study wives under 30 at enumeration in 1941 were married before their 19th birthday, and over two-thirds before their

¹¹ Whelpton and Kiser, *Ibid.*, p. 231. (Reprint, p. 11.)

Table 19. Number of children ever born per 100 wives by age and age at marriage for Protestant couples in (a) the Household Survey, (b) the original universe of couples eligible for the Study and, (c) the inflated Study sample.

AGE OF WIFE AT ENUMERATION AND AT MARRIAGE	HOUSEHOLD SURVEY	STUDY	
		Original Universe	Inflated Sample
CHILDREN BORN 100 WIVES			
Age at Enumeration: 25-29	120	226	205
Age at Marriage: Under 17	249	243	217
" " "	175	187 ^a	179 ^a
Age at Enumeration: 30-34	158	180	181
Age at Marriage: Under 17	293	236*	249*
" " "	220	191	190
" " "	151	162	164
Age at Enumeration: 35-39	199	135	136
Age at Marriage: 20-22	181	146	146
" " "	137	134	135
" " "	99	104 ^b	107 ^b
Age at Enumeration: 40-44	219	91	90*
Age at Marriage: 26-28	129	102	93*
" " "	69	58**	—
NUMBER OF WIVES			
Age at Enumeration: 25-29	7,167	253	218
Age at Marriage: Under 17	605	176	148
" " "	2,162	77 ^a	70 ^a
Age at Enumeration: 30-34	6,857	1,387	1,105
Age at Marriage: Under 17	651	42	39
" " "	1,964	755	616
" " "	1,810	590	450
Age at Enumeration: 35-39	6,038	780	571
Age at Marriage: 20-22	1,705	253	178
" " "	967	449	347
" " "	503	78 ^b	46 ^b
Age at Enumeration: 40-44	5,283	163	82
Age at Marriage: 26-28	420	125	75
" " "	179	38 ^a	7 ^a

* Based on 25-99 cases (see numbers in lower part of table).

^a Age at marriage: 17-18.

^b Age at marriage: 26-27.

^{*} Age at marriage: 29.

17th birthday. Conversely, wives 40-44 in 1941 were married at ages 25-29.⁴⁴ In general, therefore, whereas the younger women eligible for the Study married very young, the older women married comparatively late. Because of this fact, and because duration of marriage is held virtually constant, the age-specific cumulative fertility rates (based on total number of children ever born) of wives in the Study vary inversely with age of wife. In the population as a whole, of course, the total past fertility varies directly with age of wife. The situation is shown graphically in Figure 1, based on Table 19. Thus, at ages 25-29 the number of children ever born per one hundred wives is considerably higher for the Study than for the Survey. At ages 30-34 the difference is smaller but in the same direction. At ages 35-39 the order is reversed and the difference is important. At ages 40-44 the rate for the Study is less than half that for the Survey.⁴⁵

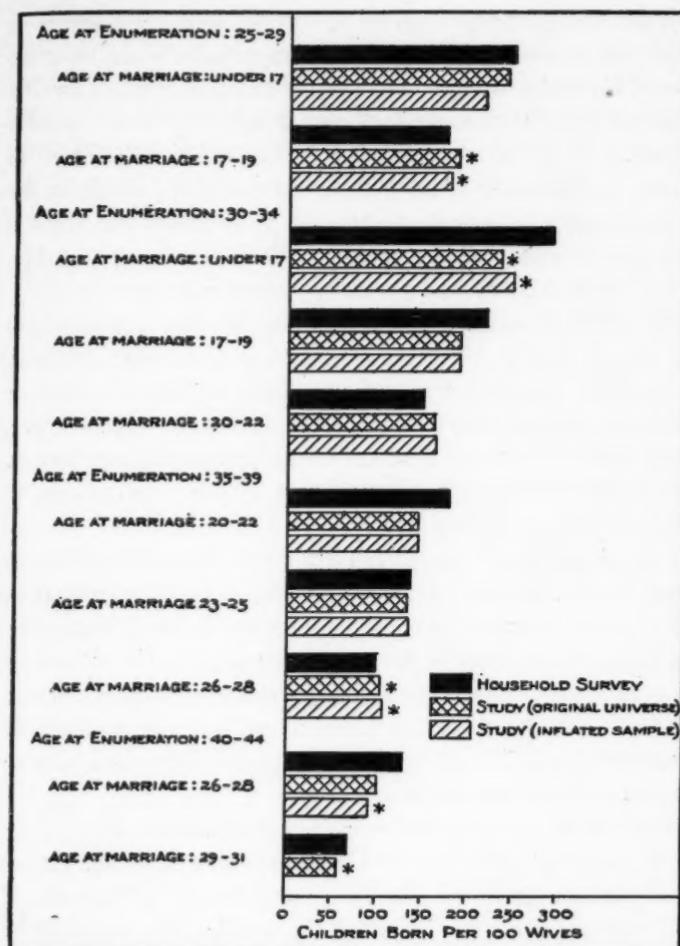
Obviously, therefore, simple age-specific comparisons are not sufficiently refined for comparisons of the Survey and Study couples with respect to fertility. Age at marriage as well as age must be considered. When this is done, as in Figure 2, the fertility rates for the Survey and Study couples become much more similar in magnitude and in pattern of variation. The differences in magnitude that do exist are rather consistently in the direction of higher rates for the Survey than for the Study couples, but this would be expected in view of the educational and urban residence restrictions of the latter group.

SUMMARY

The Household Survey of Indianapolis yielded a total of 2,589 couples qualifying for inclusion in the intensive Study of Social and Psychological Factors Affecting Fertility. Because the number

⁴⁴ Wives 30 or over at marriage were not eligible for the Study.

⁴⁵ In the above, the Survey data (like those for the Study) are restricted to Protestant couples. Age-specific rates for couples of all religions in the Survey differ very slightly because the Survey is heavily weighted by Protestants. See Whelpton and Kiser, *Ibid.*, p. 229. (Reprint, p. 9.)



* Rate based on 25-99 cases.

Fig. 2. Number of children ever born per one hundred wives by age and age at marriage for Protestant couples in (a) the Household Survey; (b) the original universe of couples eligible for the Study; and (c) the inflated Study sample. See Table 19.

of one and two-child families planned as to size was much larger

than the corresponding number of families with no children, or with three or more, it was desirable to sample the one and two-child groups. Similarly, the high proportion of childless couples classified as relatively sterile led to the secondary sampling of sterile childless couples. The primary sampling ratios adopted were designed to secure approximately equal numbers of completed schedules for fecund couples of each parity planned as to family size, couples with four or more children being combined into one group. The sample for each parity was stratified by rental value of the home.

For various reasons the interviewers were unable to fill out schedules for, or even to call upon, all of the couples assigned to them. Some of the couples had moved away from Indianapolis or to an unknown address, some were not found at home at repeated visits, others were seen but were not interviewed for miscellaneous reasons. Available tests indicate that the omission of these couples causes no substantial bias in the final results.

Approximately 11.1 per cent of the couples seen by interviewers refused to cooperate in the Study. Owing to a higher proportion childless, the average number of children per family is smaller for the uncooperative than for the cooperative couples. The differences between the two groups are negligible for the remaining items on the Survey schedules, but may be important for certain items on the detailed Study schedules. Hence, the exclusion of the uncooperative couples may have introduced some bias.

The sampling plan made necessary the adjustment of any summary rate or average computed from data obtained from couples of all parities combined. Provision for automatic adjustments of this type was made by "inflating" the sample. This was done by duplicating predetermined numbers of punch cards, selected at random, for couples of given parity. Tests indicate that the total inflated group is very similar to the original universe of eligible couples not only with respect to distribution by family size, but also with respect to various other descriptive items.

The number of children born for 100 wives by age at marriage and current age of wife is somewhat lower as a rule for couples in the Study than for other once-married Protestant couples in Indianapolis. The differences probably are due to the exclusion of couples from the Study if either the wife or husband had not completed the eighth grade, or if the couple had not lived in a large city most of the time since marriage.